PHD

Higher education and ethnicity: Exploring patterns and inequalities in ethnic minorities’ attainment in the UK universities and labour market performance

Kollydas, Kostas

Award date:
2021

Awarding institution:
University of Bath

Higher education and ethnicity: Exploring patterns and inequalities in ethnic minorities' attainment in the UK universities and labour market performance

Konstantinos Kollydas

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Economics

April 2021

Summary

The overarching aim of this research is twofold. First, I explore the disparities in academic attainment and progression between ethnic minority undergraduates and their White peers in the UK, in terms of the probability of graduating with a "good" class of degree and the dropout likelihood. Second, I estimate the size of ethnic pay inequalities in the British labour market among UK-born graduates and quantify the contribution of specific characteristics to the wage differential.

Regarding the likelihood of degree non-completion, for the first time in the ethnicity context, I distinguish between compulsory dismissal (because of academic failure) and voluntary dropout (because of personal, financial, or other reasons), acknowledging that an effective policy response to student withdrawal ought to be closely associated with the dropout causes. I provide firm evidence that all ethnic minority groups have, on average, a higher probability of failing their degrees than White students, even after allowing for differences in a wide range of socio-demographic and university characteristics. Most worryingly, Black African students are twice as likely (7.1%) as their White peers (3.8%) to fail. On the contrary, White students are more inclined towards quitting voluntarily than ethnic minorities, although the differences are smaller than those relating to academic dismissal. The probability of attaining a good degree class stands at 76.2% for White students, and the difference relative to ethnic minorities ranges from 4.3 percentage points (for students from a Mixed ethnic background) to 15.1 percentage points (for Black African students). These performance gaps remain striking across all types of institutions and social class levels, suggesting that ethnic minorities' under-attainment is a pervasive issue.

With respect to the second target of this thesis, a key conclusion of this work is that having graduated from university does not eliminate ethnic pay inequalities. For Black men, the labour market experience is more disappointing, as their wage penalties stand at higher levels (16.7%) than those of Black women (4.5%) and the other ethnic groups. The occupational segregation and the shorter job tenure of ethnic minorities relative to White people make up the primary determinants of pay gaps relating to observed characteristics. Although it is difficult to disentangle the extent of racial discrimination from other unobserved factors that also shape the wage levels, by implementing a novel method that partially addresses the selection on unobservable determinants of earnings, I find supportive evidence of ethnic discrimination, particularly against Black and Asian men. I reveal that ethnic penalties worsen with age across both genders, implying that some mechanisms endogenous to the labour market (such as discrepancies in promotion procedures, access to company-specific training, and the time required to find a job, which, in turn, affects the accumulative experience) should, at least partially, explain these results.

These research findings should help policymakers better comprehend the mechanisms that produce diverse academic and earnings outcomes and identify targeted strategies for each ethnic group. In this context, I propose certain policymaking actions that should alleviate the barriers faced by ethnic minorities in higher education and the UK labour market.

To Ioannis Kollydas

"Γιατί πιστεύοντας με πάθος κάτι που δεν υπάρχει ακόμα, το δημιουργούμε. Ανύπαρχτο είναι ό,τι δεν πεθυμήσαμε αρκετά [...]"

Νίκος Καζαντζάκης,
"Αναφορά στον Γκρέκο"

"For, by believing passionately in something which still does not exist, we create it. The nonexistent is whatever we have not sufficiently desired [...]"

Nikos Kazantzakis,
"Report to Greco"
Translated from Greek by P.A. Bien

Acknowledgements

It is hard to fully express into words my profound gratitude to my supervisor, Joanna Clifton-Sprigg, for believing in me and providing continuous encouragement, mentoring, great intuition and guidance on the thesis, but also the freedom to develop my own research ideas, and for everything she has taught me during my studies. Her open nature and out-of-the-box thinking comprise a rare combination, which has helped me improve as a researcher.

I am indebted to Matt Dickson for his invaluable support on various aspects of my PhD, the useful comments on this thesis, the opportunities he has opened doors to, and our really stimulating discussions. My compliments also go to my interim supervisor, Kerry Papps, for all his contributions to this journey — I have been very fortunate to know them.

I am deeply grateful to John T.E. Richardson for his insightful feedback on the thesis and our inspirational conversations about literally everything. I also wish to thank my external examiner, Simon Burgess, and Rebekka Christopoulou, Judith Delaney, Jonathan James, Štěpán Jurajda, Chris Martin, Thanos Mergoupis, Nikos Kokonas, Javier Valbuena, Jan van Ours, Lorraine Wong, and Yu Zhu for their useful comments on different parts of the present work. Moreover, it has been a great pleasure to work with Alistair Hunt, Maik Schneider, Samuel Bodunrin, Yidi Sun, Imran Shah, Van Phan, and Ian Corrick on different projects or teaching.

Throughout my PhD years, I have met many interesting people and excellent colleagues with whom I have shared quality time, ideas, and concerns. I am grateful to all of them — special thanks to Yuexian, Magdalyn, Jo, Richard, Cristina, Peter, and Lingsi.

Above all, I have no words to thank enough my family (Andriani, Dina, Elli, Nikos, Odysseas) for being there for me and inspiring me every day. If one thing is certain, it is that without their love and unconditional support, I would not have made this all possible.

Many thanks also go to all close friends — especially Panagiotis Mallios, Vasilis Tsakalis, the Zafeiris brothers, George Loukopoulos, and Panagiotis Despotis — for their precious companion.

Finally, I would like to sincerely thank the University of Bath (Widening Participation fund) for financially supporting my research in the form of a PhD studentship.

Contents

List of Tables

List of Figures

List of Abbreviations

AAP	Average Adjusted Prediction
AME	Average Marginal Effect
APS	Annual Population Survey
ATT	Average Treatment Effect on the Treated
BAME	Black, Asian, and Minority Ethnic
BIS	Department for Business, Innovation and Skills
CIA	Conditional Independence Assumption
DfE	Department for Education
EB	Entropy Balancing
FE	Further Education
GCE	General Certificate of Education
GCSE	General Certificate of Secondary Education
GDP	Gross Domestic Product
GPA	Grade Point Average
HE	Higher Education
HEFCE	Higher Education Funding Council for England
HESA	Higher Education Statistics Agency
HM	Hausman-McFadden
IIA	Independence of Irrelevant Alternatives
ILO	International Labour Organisation
IPWRA	Inverse Probability Weighted Regression Adjustment
JACS	Joint Academic Coding System
LAD	Least Absolute Deviations
LEA	Local Education Authority
LEM	Law, Economics, and Management
LFS	Labour Force Survey
LP	Linear Programming
LPM	Linear Probability Model
LR	Likelihood Ratio
MD	Mahalanobis Distance
ME	Marginal Effect
MEM	Marginal Effect at the Means
MLR	Multinomial Logistic Regression
NAO	National Audit Office
NCIHE	National Committee of Inquiry into Higher Education
NUS	National Union of Students
OECD	Organisation for Economic Co-operation and Development
OfS	Office for Students
OLS	Ordinary Least Squares
ONS	Office for National Statistics
PSM	Propensity Score Matching
QR	Quantile Regression
REF	Research Excellence Framework
RG	Russell Group
RPI	Retail Price Index
SE	Standard Errors
SMF	Social Market Foundation
STEM	Science, Technology, Engineering, and Mathematics
TEF	Teaching Excellence Framework
UCAS	Universities and Colleges Admissions Service
UUK	Universities UK

1. Introduction

1.1 Motivation and primary research objectives

Ethnic disparities in higher education and labour market outcomes have been at the heart of government and social debate over the last years in the UK (Bulman, 2017; Khan, 2020; Almeida, Brodnock and Lordan, 2021). In an adjacent context, the "Black Lives Matter" movement and the disproportionate ethnic effects of the ongoing COVID-19 pandemic on earnings and employment (Bracke et al., 2021) have further attracted the attention of the media, academics, and society as a whole. A recent Government's response to the disappointing figures around specific ethnic minorities' academic performance concerning dropout rates and achievement of good grades at university was to apply additional pressure on universities to report attainment information analysed by ethnicity and socio-economic background (DfE, 2019). For instance, the official data show that the proportion of Black undergraduates no longer in higher education the academic year following entry was 10.3% in 2014/15, while the corresponding share of White, Indian, and Chinese students was significantly lower, standing at 6.9%, 5.4%, and 4.0%, respectively (OfS, 2018).

Moreover, ethnic pay inequalities seem to persist in the UK labour market despite the existing laws against racial discrimination in employment (e.g., Equality Act, 2010). For example, the median hourly earnings in England and Wales for Pakistani (£10.55), Bangladeshi (£10.58) and Black African (£11.50) people were remarkably lower compared to their White British (£12.49) counterparts in 2019, whereas some ethnic groups (such as Chinese and Indian) earn more than the White British employees (ONS, 2020). In this respect, in 2018, the former Prime Minister, Theresa May, considered introducing a rule that would oblige UK companies to publish figures pertaining to "ethnicity wage gaps" (Swinford, 2018).

However, merely looking at raw figures provides an incomplete picture of ethnic inequalities in higher education and the labour market, especially if one does not consider how ethnicity and other critical characteristics that determine academic performance and earnings interact with each other. Previous research (e.g., Crawford and Greaves, 2015) has analysed the notable progress in ethnic minorities' fair access to universities (UCAS, 2020). Nevertheless, the literature on the economics of education has placed less focus on examining the extent and determinants of ethnic differences in the probability of non-completion (Vignoles and Powdthavee, 2009). In particular, none of the studies to date exploring ethnic disparities in student attrition in the UK has differentiated between involuntary withdrawal (because of academic failure) and voluntary dropout (because of personal, financial, or other reasons). Neglecting to implement such distinction has substantial implications from a policymaking viewpoint, as it prevents identifying target groups of students requiring specific forms of support.

Secondly, previous research investigating disparities in the probability of getting a "good" degree class is based on relatively old data with inadequate information about some crucial factors that influence academic attainment, such as prior educational achievement and students' socio-economic classification (e.g., Broecke and Nicholls, 2007). Furthermore, most of the earlier works rarely examine the institutional-level variation in academic performance and neglect the significant role of interdependence in explaining the "ethnic attainment gaps" in higher education (that is, the difference in academic attainment between each ethnic minority group and their White peers). Instead, they assume that the effect of ethnicity on the likelihood of attaining a good degree is additive rather than operating interactively. Given that a good degree class is associated with better earnings prospects later in life (Naylor, Smith and Telhaj, 2016; Feng and Graetz, 2017), it is essential to acquire more comprehensive knowledge of the academic performance correlates and mechanisms that compromise achievement among ethnic minorities.

Thirdly, very little is known about whether ethnic pay inequalities exist among UK-born higher education graduates in the British labour market (Zwysen and Longhi, 2018), and there is barely any evidence of how these wage gaps develop over the life course. This is important to explore because the proportion of the (highly qualified) university graduates in the UK workforce has proliferated over the last three decades in line with the expansion of the higher education sector. By focusing on UK-born graduates, the present research attempts to strengthen our understanding of the unexplained component of wage gaps (which is related to unobserved characteristics and discrimination), thus directing policymaking to targeted interventions.

Motivated by the limited knowledge regarding the matters mentioned above and the related Government's actions, I seek to address these literature gaps and discuss the relevant policy implications. The first empirical paper of this work examines the effect of ethnicity on the likelihood of dropping out of UK universities. By using recent personal-level data from the Higher Education Statistics Agency (HESA) for the academic years 2010/11-2014/15, I estimate the ethnic differences in the likelihood of graduation, voluntary dropout, and involuntary withdrawal, conditional on an extensive range of characteristics (such as age, gender, socio-economic background, parental education, disability status, region and type of university, institution quality measures, subject of study, mode of study, peer effects, and prior educational ability). In addition, I explore how the effect of ethnicity on the likelihood of academic failure varies according to university type and students' gender, socio-economic background, and prior educational attainment.

The second empirical paper exploits the census nature of HESA's data for the same academic years (2010/11-2014/15) to estimate the probability of obtaining a "good" bachelor's degree (that is, a first-class or upper-second class of degree), contingent on students completing their studies. I rely on a detailed ethnicity classification (eleven ethnic groups), which contributes to a better explanation of the variation in the dependent variable. Similarly to the first paper, I decipher the heterogeneous effects of ethnicity on the propensity of earning a good degree across genders, institution types, and students' social classes. The first two papers focus on young undergraduate students (aged under 21 on entry) who lived in the UK before commencing their degree course.

The third empirical paper draws on restricted-access data from the UK Annual Population Survey (APS) for years 2013-2018 to explore the magnitude of wage differences between UK-born university graduates from Black, Asian, or Mixed/Other minority backgrounds and their White counterparts. Moreover, I scrutinise the heterogeneous effects of ethnicity on wages based on the type of university attended, subject area of study, degree class, major occupation group, workplace size, and across age bands. For the first time in the ethnicity framework, I evaluate the robustness of the regression results by employing a novel partial identification approach (Oster, 2017), which calculates estimates correcting for selection on unobserved characteristics. In doing so, I provide supportive evidence of ethnic discrimination against specific ethnic groups. By adopting a decomposition method, I also quantify the contribution of each main group of observable characteristics to the earnings differential between White and non-White employees.

The rest of this introductory chapter reports the key findings of the present research (section 1.2), followed by some background information and a brief description of the UK education system (1.3) and the thesis structure outline (1.4).

1.2 Main results

In the first paper, using multinomial logistic regression models, I find that all ethnic minority groups are on average more likely to leave university because of academic failure than White undergraduates. The picture is more worrying for Black African students, whose probability of failing academically (7.1%) is nearly two times higher than that of their White peers (3.8%). The ethnic gaps in academic failure are more extensive for men than women across most ethnic minority groups and are less noticeable in the Russell Group institutions. Also, the ethnic differences in involuntary dropout do not deviate significantly across various levels of students' socio-economic classification. In contrast, White undergraduates exhibit a greater propensity to withdraw voluntarily (for example, due to personal reasons or because they choose to go into employment) than all ethnic minorities. Taking the above figures together, the first-degree completion rates conditional on the complete set of observed characteristics are approximately one percentage point higher for White undergraduates (89.4%) relative to ethnic minorities, except for Indian (89.6%) and Chinese (89.5%) students.

The logistic regression results from the second empirical paper reveal that the average adjusted predicted probability of graduating with a good degree is significantly higher for White students (76.2%) than for all ethnic minority groups, and the picture is shocking for Black minorities. Specifically, the ethnic gaps in attainment range from 15.1 percentage points for Black African students to 4.3 percentage points for the Mixed ethnic group (relative to similarly situated White students). These ethnic disparities remain impressively robust even after performing specific sensitivity checks, such as incorporating university and course fixed effects into the analysis. The ethnic gaps in the likelihood of obtaining a good degree are larger amongst women than men across all Asian ethnic groups (Indian, Pakistani, Bangladeshi, and other Asian), except for Chinese, while the ethnic differences in attainment remain statistically significant within all university types.

The OLS regression results from the third empirical paper establish stark ethnic penalties (in terms of hourly wages) of 16.7% for UK-born Black male degree holders relative to equally qualified White employees. The wage gaps are considerably smaller for Asian men (-4.1%) and statistically insignificant (-1.3%) for male employees from Mixed/Other ethnic backgrounds. Black and Asian women experience substantially lower ethnic pay differences than men, standing at -4.5% and -2.0%, respectively. An important finding of this paper is that ethnic penalties among UK-born graduates exacerbate with age. The overall picture is particularly concerning for the Black community, as the average life-course ethnic penalties for Black men remain unequivocally robust even after applying Oster's (2017) identification method, implying racial discrimination. Similarly, the OLS results are robust to various matching techniques that adjust for the selection on observable characteristics. Finally, by decomposing the pay differential into explained (30% of the total wage gap) and unexplained (70%) components, I find that imbalances in job characteristics (such as the under-representation of ethnic minorities in highly paid occupations and their shorter job tenure) account for half of the total wage differential between White and non-White employees.

1.3 Background

1.3.1 Preface

Ethnicity represents a social organisation concept and is often used to classify people according to whether they exhibit a shared perception of specific attributes (such as language, religious connections, history, customs and beliefs, cultural distinctiveness, geographical origin, and race). In sociological literature, there has been a debate that ethnicity is a social process outcome that develops and transforms over time rather than merely a characteristic obtained through birth (see a discussion in Wimmer, 2008). Ethnicity is considered more subjective than race, which is closely related to skin colour and nationality.

In the decades preceding the 1991 Census, individuals living in the UK were categorised based on which Commonwealth country they were born in (in the 1960s), their White or "coloured" appearance (in the 1970s), or broad national groups (1980s) (Laux, 2019). For the first time, the 1991 Census introduced an ethnicity question on a self-identification basis, which many Government surveys and research studies subsequently utilised. The following Censuses adjusted this question and updated the ethnic categories classification to adapt to the drastically changing and ethnically diverse UK population. Specifically, the proportion of non-White people in England and Wales has increased from 7% in 1991, to 9% in 2001, 14% in 2011, and 15% in 2016 (Jivraj, 2012; ONS, 2019).

Of the total 2011 population in England and Wales (56 million), 86% identified as “White”, 8% as “Asian/Asian British”, 3% as “Black/African/Caribbean/Black British”, and 3% as “Mixed” or “Other” ethnic groups (ONS, 2012). The population projections estimate that the share of ethnic minority people will reach 27% in 2051 (Khan, 2020). Hence, the ethnic inequalities documented in the present research affect increasingly more people. Therefore, there is a need to understand and tackle the impediments driving the ethnic disparities in higher education performance and the labour market.

In the present work, I consider ethnicity a self-assigned individual characteristic adopted to explain the different university and earnings outcomes in the UK. I use the terms “ethnic gaps” and “ethnic penalties” to describe the inequalities in the outcomes of interest (such as academic attainment and wages) between a specific ethnic minority group and the reference White group. The White category includes people from White British, Irish, Gypsy or Irish Traveller, and any other White backgrounds (such as White Europeans and Australians), thus containing some White ethnic minorities. However, in line with the practice adopted by other studies in the field (e.g., Richardson, Mittelmeier and Rienties, 2020), the “ethnic minority” term used throughout this research (though strictly inaccurate) considers only non-White individuals.

1.3.2 The education system in the UK

The UK education system includes five broad phases: “early childhood”, “primary education”, “secondary education”, “further education”, and “higher education”. Schooling is compulsory for all pupils aged 5-16. As I discuss below, England recently extended the mandatory education or training until students turn 18.

1.3.2.1 Compulsory schooling

Children aged 4-11 normally attend primary schools. When they turn 11, they usually move to secondary school immediately, although in some cases the progression is made through “middle schools”, which cover pupils aged 8-13. Secondary education comprises different school types, such as community schools, academies, free schools, and city technology colleges. Institutions can be state schools (which are free for all children) or independent (private) schools (which charge fees). As of 2020, the total number of schools in England (nursery, primary and secondary) was 24.4 thousand, 9.6% of which were private (DfE, 2020). The number of all pupils reached 8.9 million in 2020 (6.5% of whom went to independent schools).

A significant characteristic of England's primary and secondary education sector is the homogeneous students' assessment system (Dustmann, Machin and Schönberg, 2010). The national curriculum, which defines the programmes and performance objectives across all phases, includes 4 "Key Stages" and 12 subjects (DfE, 2014). Primary schools involve "Key Stage 1" (pupils aged 5 to 7) and "Key Stage 2" (ages 7-11), while secondary education is associated with "Key Stage 3" (ages 11-14) and "Key Stage 4" (ages 14-16). After finishing Key Stage 4 (year 11), which signifies the completion of the mandatory schooling, pupils sit for the "General Certificate of Secondary Education" (GCSE) state examinations. Students can choose several GCSE exams in various subjects, although English, Maths and Sciences are compulsory.

1.3.2.2 Post-16 destinations

The so-called "raising the participation age" (RPA) constitutes a recent revision to England's education system, obliging all students to remain in education or training until they turn 18 (DfE, 2016). Specifically, at the age of 16, students living in England must choose among three options: a) they can continue in full-time education (by attending, for instance, a college or a secondary school with an attached sixth form); b) they can commence an apprenticeship or traineeship; or c) they can work or volunteer for at least twenty hours a week whilst studying part-time. Pupils in Scotland, Wales, and Northern Ireland can leave compulsory schooling after they turn 16.

While compulsory schooling is relatively homogeneous across programmes, subjects of study and examinations, students can follow many different routes after age 16 (Hupkau et al., 2017). In particular, the vast number of academic, vocational, and other qualifications compose a diverse landscape. Figure 1.1 outlines the general structure of the post-16 education system in England and the possible routes students can take to access university. The variations in the education systems across the UK countries are relatively small (Brennan, 2020). Further education (FE) is essentially offered by colleges (such as "general FE and tertiary colleges", "sixth form colleges", "specialist colleges") and community learning establishments. It differs from the more advanced degree-level higher education provided by the universities. FE typically includes provision ranging from elementary skills training to higher vocational qualifications. Sixth form schools may also offer specific FE courses.

Students pursuing the “academic track” usually take advanced (A) level exams for the “General Certificate of Education” (GCE). GCE A-levels are countrywide-examined academic tests taken at 18 (at the end of “Key Stage 5”). Many UK universities recognise A-levels as the typical prerequisite for evaluating applicants’ suitability and quality. Notwithstanding, an increased share of students access HE through other qualification types (such as vocational qualifications) or mixed qualification routes. Specifically, UCAS (2016) provides a description of different pathways into higher education, classified as “academic”, “applied general”, “technical”, and “occupational”.

1.3.2.3 Higher education and university admissions process

The UK HE system has grown enormously over the last fifty years. The student numbers rose from 400 thousand in the 1960s to 2 million in the early 2000s and 2.38 million in the 2018/19 academic year (Blanden and Machin, 2004; Universities UK, 2020). The HE participation rates witnessed a considerable increase after the supply-side growth in the 1990s. The ending of the “binary divide” between universities and polytechnics (as a consequence of the “1992 Further and Higher Education Act”) caused substantial structural changes to the UK HE system. As a corollary, over 40 former polytechnics were accorded degree-awarding powers. In addition, the university types and the available courses have also expanded, as students now have more choices when entering HE, considering the differentiation of the HE provision (Croxford and Raffe, 2014).

Students aiming to go to university apply via the “Universities and Colleges Admissions Service” (UCAS), a centralised admissions procedure. UCAS covers the whole of the UK and is used by most young applicants, who typically apply during the final year of school or after having completed school. Each university’s admission staff decides whether an application is successful. Universities offer a place to students contingent on their academic performance (such as the A-level achievements). If the exam results are unknown during the application, university decisions rely on predicted grades provided by students’ teachers and enrolments are subject to the applicants attaining specific grades in the respective assessments. Students commonly apply for similar subjects at different universities and submit their applications by choosing up to five courses in total.

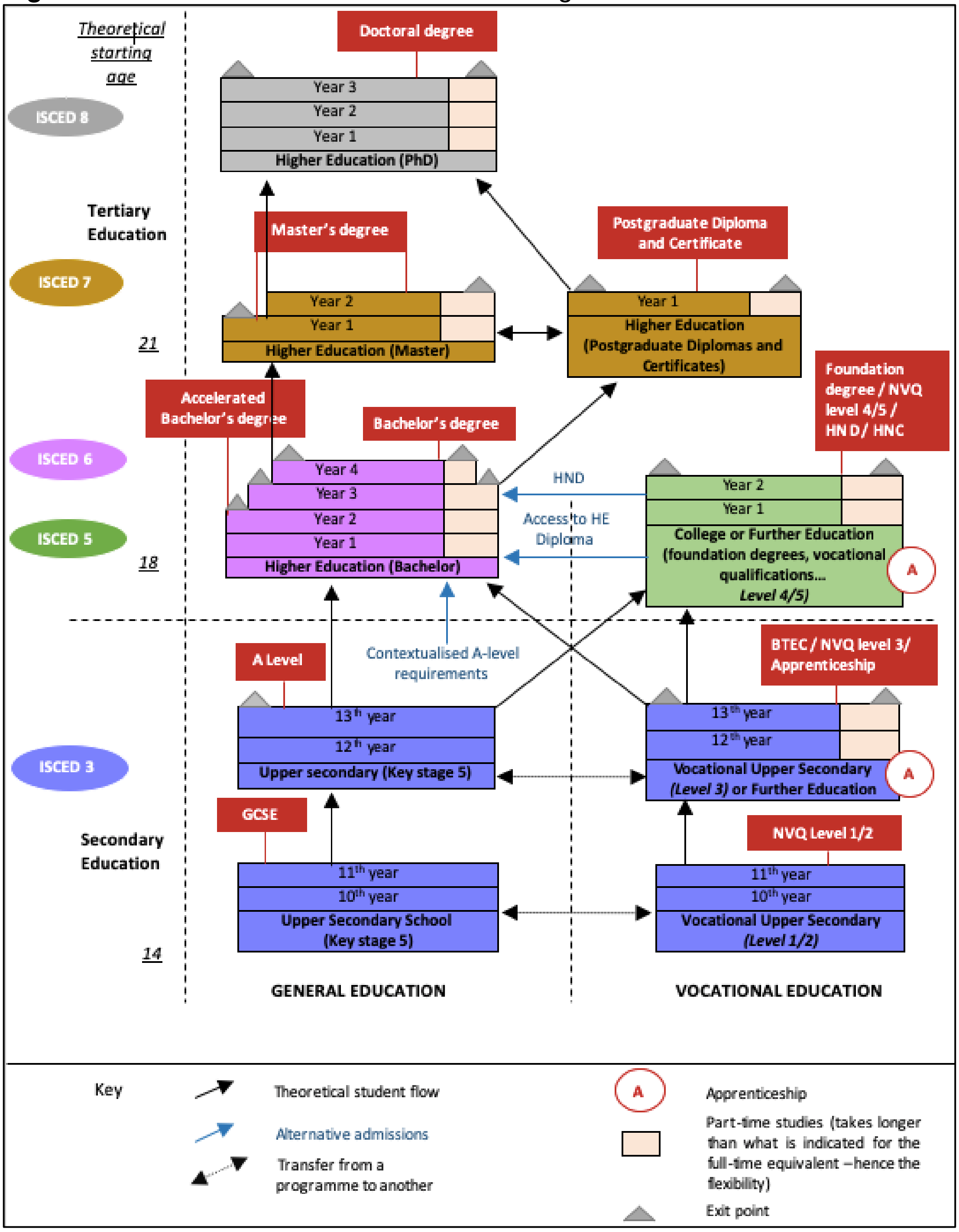
Passing three A-levels with specific grades (or holding equivalent qualifications) is the standard requirement to enter a UK university, although this varies across institutions and courses. Other requirements relate to student’s suitability (such as interests, talents, and experience) revealed in the written statement by the applicant and, less commonly, admissions tests or interviews (particularly for prestigious universities, such as Oxford or Cambridge).

UCAS data reveal that the HE 18-year-old entry rates[1] increased substantially during the 2007-2017 period in England, although remarkable differences persist among ethnic groups (UCAS, 2017). Chinese people have consistently seen the highest access rates (2017: 63.0%, up by ten percentage points relative to 2007), whereas White students witness the lowest rates (29.3% in 2017 compared to 22.2% ten years earlier). The Mixed ethnic group displayed the second lowest entry rate (34% in 2017), while the corresponding figure for Black applicants was 40.4% in 2017, showing a notable rise since 2007 (23.1%).

Students can also apply through the “Clearing process” (which relates to the remaining vacancies in the courses) if they did not fulfil the university requirements or apply after the 30th of June for undergraduate programmes commencing the following academic year. Previous evidence suggests that ethnic minority candidates are more likely than White students to access university through clearing (Connor et al., 2004).

[1] *According to UCAS, the entry rates represent “the number of acceptances from an application cycle divided by the estimated base population”. Figures presented here refer to full-time undergraduate studies for 18-year-olds living in England.*

Figure 1.1. Post-16 education structure in England

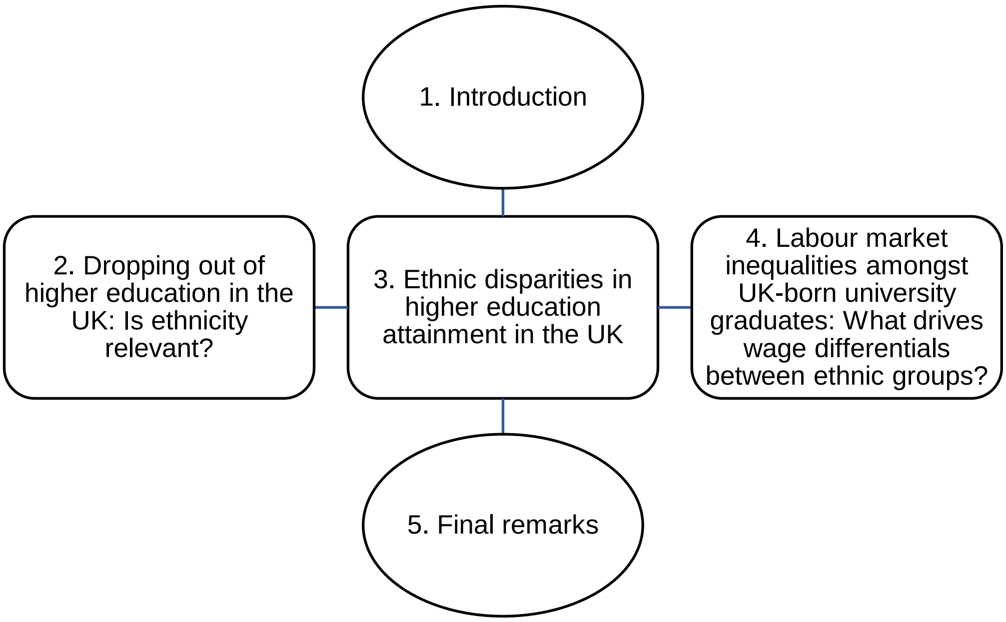


Note: The education levels are based on the "International Standard Classification of Education" (ISCED). GCSE: "General Certificate of Secondary Education"; NVQ: "National Vocational Qualification"; BTEC: "Business and Technology Education Council" qualification; HND: "Higher National Diploma"; HNC: "Higher National Certificate".
Source: Figure taken from Brennan (2020).

1.4 Research structure

The rest of the present dissertation comprises the three separate, but interlinked, empirical papers described earlier and the conclusion chapter (see Figure 1.2).

Figure 1.2. Thesis structure



References of Chapter 1

Almeida, T., Brodnock, E. and Lordan, G., 2021. *Black women are missing in the UK's top 1%* [Online]. London School of Economics and Political Science. Available from: https://blogs.lse.ac.uk/businessreview/2021/03/03/black-women-are-missing-in-the-uks-top-1/ [Accessed 15 March 2021].

Blanden, J. and Machin, S., 2004. Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy,* 51(2), pp. 230-249.

Bracke, P., Croxson, K., Leary, J. and Wood, J., 2021. Covid-19 and the UK's BAME communities – an economic perspective. *Financial Conduct Authority* [Online]. Available from: https://www.fca.org.uk/insight/covid-19-and-uk-bame-communities-economic-perspective [Accessed 20 February 2021].

Brennan, J., 2020. *Flexible Learning Pathways in British Higher Education: A Decentralized and Market-Based System.* International Institute for Educational Planning (UNESCO). Available from: https://www.qaa.ac.uk/docs/qaa/about-us/flexible-learning-pathways.pdf [Accessed 20 December 2020].

Broecke, S. and Nicholls, T., 2007. *Ethnicity and degree attainment.* London: Department for Education and Skills, (Research Report RW92).

Bulman, M., 2017. Black students 50% more likely to drop out of university, new figures reveal. *The Independent* [Online]. Available from: https://www.independent.co.uk/news/uk/home-news/black-students-drop-out-university-figures-a7847731.html [Accessed 20 December 2020].

Connor, H., Tyers, C., Modood, T. and Hillage, J., 2004. *Why the difference? A closer look at higher education minority ethnic students and graduates.* Nottingham: DfES (Department for Education and Skills), (RR552).

Crawford, C. and Greaves, E., 2015. *Socio-economic, ethnic and gender differences in HE participation.* London: BIS (Department for Business Innovation and Skills), (186).

Croxford, L. and Raffe, D., 2014. Social class, ethnicity and access to higher education in the four countries of the UK: 1996–2010. *International Journal of Lifelong Education,* 33(1), pp. 77-95.

DfE, 2014. *The national curriculum in England.* London: DfE (Department for Education).

DfE, 2016. *Participation of young people in education, employment or training.* London: DfE (Department for Education), (DFE-00255-2016).

DfE, 2019. Universities must do more to tackle ethnic disparity. *Department for Education, The Cabinet Office* [Online]. Available from: https://www.gov.uk/government/news/universities-must-do-more-to-tackle-ethnic-disparity [Accessed 20 December 2020].

DfE, 2020. *Schools, pupils and their characteristics* [Online]. London: DfE (Department for Education). Available from: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics [Accessed 20 February 2021].

Dustmann, C., Machin, S. and Schönberg, U., 2010. Ethnicity and Educational Achievement in Compulsory Schooling. *Economic Journal,* 120(546), pp. F272-F297.

Equality Act, 2010. c. 15. *Retrieved from the UK National Archives website: http://www.legislation.gov.uk/ukpga/2010/15/contents.* London: HMSO.

Feng, A. and Graetz, G., 2017. A question of degree: The effects of degree class on labor market outcomes. *Economics of Education Review,* 61, pp. 140-161.

Hupkau, C., McNally, S., Ruiz-Valenzuela, J. and Ventura, G., 2017. Post-Compulsory Education in England: Choices and Implications. *National Institute Economic Review,* 240(1), pp. R42-R57.

Jivraj, S., 2012. *How has ethnic diversity grown 1991-2001-2011?* Manchester: ESRC Centre on Dynamics of Ethnicity (CoDE). Available from: https://www.basw.co.uk/system/files/resources/basw_44730-5_0.pdf [Accessed 20 December 2020].

Khan, O., 2020. *The Colour of Money -How racial inequalities obstruct a fair and resilient economy.* London: Runnymede. [Accessed 20 December 2020].

Laux, R., 2019. *50 years of collecting ethnicity data* [Online]. Cabinet Office. Available from: https://history.blog.gov.uk/2019/03/07/50-years-of-collecting-ethnicity-data/ [Accessed 20 December 2020].

Naylor, R., Smith, J. and Telhaj, S., 2016. Graduate returns, degree class premia and higher education expansion in the UK. *Oxford Economic Papers,* 68(2), pp. 525-545.

OfS, 2018. *Topic briefing: Black and minority ethnic (BME) students.* OfS (Office for Students). Available from: https://www.officeforstudents.org.uk/media/145556db-8183-40b8-b7af-741bf2b55d79/topic-briefing_bme-students.pdf [Accessed 20 December 2020].

ONS, 2012. *Ethnicity and National Identity in England and Wales: 2011.* Newport: ONS (Office for National Statistics).

ONS, 2019. *Population characteristics research tables.* ONS (Office for National Statistics). Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationcharacteristicsresearchtables [Accessed 20 December 2020].

ONS, 2020. *Ethnicity pay gaps: 2019* [Online]. ONS (Office for National Statistics). Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/ethnicitypaygapsingreatbritain/2019 [Accessed 20 December 2020].

Oster, E., 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics,* 37(2), pp. 187-204.

Richardson, J.T.E., Mittelmeier, J. and Rienties, B., 2020. The role of gender, social class and ethnicity in participation and academic attainment in UK higher education: an update. *Oxford Review of Education,* 46(3), pp. 346-362.

Swinford, S., 2018. All large employers will be forced to publish their 'ethnicity pay gap' under Government plans. *The Telegraph* [Online]. Available from: https://www.telegraph.co.uk/politics/2018/10/10/large-employers-will-forced-publish-ethnicity-pay-gap-government/ [Accessed 20 December 2020].

UCAS, 2016. *Progression Pathways.* Gloucestershire: UCAS (Universities and Colleges Admissions Service).

UCAS, 2017. *2017 End of Cycle Report: Patterns by applicant characteristics.* Gloucestershire: UCAS (Universities and Colleges Admissions Service).

UCAS, 2020. *2019 End of Cycle Report: Equality in England.* Cheltenham: UCAS (Universities and Colleges Admissions Service), (MD-5927).

Universities UK, 2020. *Higher education in numbers* [Online]. London: Universities UK. Available from: http://www.universitiesuk.ac.uk/facts-and-stats/Pages/higher-education-data.aspx [Accessed 20 December 2020].

Vignoles, A. and Powdthavee, N., 2009. The Socioeconomic Gap in University Dropouts. *The B.E. Journal of Economic Analysis & Policy,* 9(1).

Wimmer, A., 2008. The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory. *American Journal of Sociology,* 113(4), pp. 970-1022.

Zwysen, W. and Longhi, S., 2018. Employment and earning differences in the early career of ethnic minority British graduates: the importance of university career, parental background and area characteristics. *Journal of Ethnic and Migration Studies,* 44(1), pp. 154-172.

Statement of Authorship

This declaration concerns the article entitled:	
Dropping out of higher education in the UK: Is ethnicity relevant?	
Publication status (tick one)	
Draft manuscript ✓ **Submitted** ☐ **In review** ☐ **Accepted** ☐ **Published** ☐	
Publication details (reference)	
Copyright status (tick the appropriate statement)	
I hold the copyright for this material ✓	Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate contributed to / considerably contributed to / predominantly executed the… Formulation of ideas: 95% Design of methodology: 100% Experimental work: N/A Presentation of data in journal format:100%
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.
Signed	Konstantinos Kollydas — **Date** 24/04/2021

2. Dropping out of higher education in the UK: Is ethnicity relevant?

Abstract

Dropping out of university induces economic costs to students, universities, and society. The UK Government has recently taken actions to diminish ethnic inequalities in non-completion rates. This study exploits extensive data of over one million first-degree students for the academic years 2010/11-2014/15 to explore how the university dropout behaviour varies across ethnic groups in the UK. For the first time in the ethnicity context, this paper distinguishes between compulsory withdrawal (because of academic failure) and voluntary attrition, acknowledging that the policy response to student withdrawal should be different depending on the dropout causes. The model results provide firm evidence that, on average, all ethnic minority groups (especially Black undergraduates) have a higher probability of academic failure than White students, while the latter are the most likely to withdraw voluntarily. Unlike studies to date, this work estimates interactive effects to show that the ethnic gaps in academic failure are more pronounced for men than women and smaller in the Russell Group universities relative to other institution types. These findings help better understand the mechanisms that produce diverse dropout outcomes and identify targeted policymaking strategies for each ethnic group.

Keywords: university dropout, academic failure, ethnicity, minorities, UK higher education, HESA

JEL classification: I23, I24, J15

2.1 Introduction

Over the last decades, student attrition in the UK higher education (HE) has raised substantial political and societal concerns and provoked policy changes (Longden, 2012). Reducing non-continuation is critical from an economic perspective, as dropping out of higher education is associated with increased costs for students, universities, and society as a whole (Yorke, 1998a). In a similar vein, people who attain a degree may reap a post-university wage premium relative to those who do not complete their studies, even if the latter have the same years of education. This theory is known as the “sheepskin effect” in economic literature (Hungerford and Solon, 1987; Jaeger and Page, 1996).

This study is partly motivated by the Government’s increased efforts to reduce racial disparities in higher education by exerting pressure on universities to report data on ethnic minorities’ attainment and progression. In this context, in February 2019, the former Minister for Universities, Chris Skidmore, stated: “It cannot be right that ethnic minority students are disproportionately dropping out of university and I want to do more to focus on student experience to help ethnic minority students succeed at university” (DfE, 2019). The official raw data show that the percentage of Black undergraduates no longer in higher education the academic year following entry stood at 10.3% in 2014/15, while the corresponding proportion of White, Indian, and Chinese students was 6.9%, 5.4%, and 4.0%, respectively (OfS, 2018). However, solely focusing on raw figures and early dropouts provides only a partial picture, especially if one does not consider interactions between ethnicity and other individual and university characteristics.

There has been significant progress in ethnic minorities’ access to universities (UCAS, 2020), which has been examined by earlier studies (e.g., Crawford and Greaves, 2015). However, the research on the economics of education has placed less focus on exploring the extent and determinants of ethnic disparities in the probability of non-completion. Most importantly, none of the studies to date investigating ethnic disparities in student attrition in the UK has distinguished between involuntary withdrawal (because of academic failure) and voluntary dropout (because of personal, financial, or other reasons). Neglecting to apply such distinction can have significant implications from a policymaking standpoint, as this would prevent detecting target groups of students that require specific forms of support. Moreover, there is limited applied work centring on recent cohorts of students to analyse ethnic differences in dropout rates in the UK, particularly after controlling for university factors and students' background characteristics (Vignoles and Powdthavee, 2009). In addition, very little is known about whether ethnic differences in non-completion vary depending on the type of institutions and students' socio-economic classification.

I address these gaps in the literature by examining the effect of ethnicity on the dropout propensity, contingent on students accessing higher education. Firstly, by using recent individual-level data from the Higher Education Statistics Agency (HESA) for the academic years 2010/11-2014/15, I descriptively explore the key factors influencing the probability of first-degree completion. Secondly, I estimate the ethnic differences in the likelihood of graduation, voluntary dropout, and involuntary withdrawal, conditional on a wide range of characteristics (such as age, gender, socio-economic background, parental education, disability status, region and type of university, institution quality measures, subject of study, mode of study, peer effects, and prior educational ability). For each ethnic group separately, I also present the predicted probabilities of the three outcomes mentioned above. Thirdly, I scrutinise the heterogeneous effects of ethnicity on the propensity of academic failure according to gender, the type of institution attended, and the students' socio-economic background.

Several studies from the education and economic literature in the UK have examined the principal factors that influence dropout rates in higher education across universities and subjects of study (Johnes, 1990, 1997; Smith and Naylor, 2002; Christie, Munro and Fisher, 2004; Johnes and McNabb, 2004; Rose-Adams, 2013; Petrie and Keohane, 2019). For example, the research identifies sizeable differences in dropout rates based on students' family background (Yorke and Thomas, 2003; Crawford, 2014a). Notwithstanding, only a few works have concentrated on ethnicity's role in student retention (e.g., Singh, 1990; National Audit Office, 2007; Vignoles and Powdthavee, 2009). Although some studies (e.g., Yorke and Longden, 2008) control for broad ethnic groups in their analyses (such as White versus non-White students), this ethnic generalisation masks possible inter-group dissimilarities.

I make three contributions to the extant literature. First, by employing multinomial logit models, I distinguish between voluntary and involuntary withdrawal when estimating the ethnic disparities in the probability of dropping out of higher education. Although Johnes and McNabb (2004) were the first to apply this approach in the UK context, they did not include ethnicity as a control variable in their analysis, possibly because of data limitations. A second contribution of the present study is to classify ethnicity into eleven groups. Apart from the fact that this detailed categorisation better explains the variation in the likelihood of graduating or leaving university, it should also allow higher education institutions and decision-makers to implement targeted interventions and forms of assistance to specific ethnic minorities. This paper's third contribution is to formally investigate interaction effects between ethnicity and students' gender, socio-economic background, type of university attended, and prior educational attainment. Nevertheless, interpreting the interaction effect between two independent variables in non-linear regressions should not hinge on the statistical significance, sign, and coefficient size of the respective product term (Ai and Norton, 2003). For this reason, I adopt a recently emerged method (Buis, 2010; Long and Mustillo, 2018; Mize, 2019), which quantifies the impact of the intersectional relationships by testing whether the marginal effects (based on the predicted probabilities) are equal across different levels of the interacted regressors.

This paper focuses on students who lived in the UK before commencing their degree course. By focusing on this group, I minimise the confounding effects of English fluency and other unobserved characteristics in the HESA datasets that differ between international and UK-domiciled students and may disproportionately influence the likelihood of degree completion (Johnes and McNabb, 2004). To estimate the probability of non-completion, I rely on the cohorts of first-degree students who leave university in the same academic year (because they complete their course or because they fail/withdraw), regardless of when they commenced their studies. Hence, this approach is based on "exit cohorts" (Johnes, 1997; Lee and Buckthorpe, 2008) and has the advantage that it does not require following through the students for many years in university. Also, unlike the alternative "continuation method", which calculates the proportion of undergraduate entrants continuing their study from the first year to the second one, the approach used here allows students to drop out at any time of their academic life. I concentrate on UK-domiciled young students who are aged under 21 on entry. The young students represent the vast majority (more than 75%) of the sample of people attending a first-degree programme in the current analysis period. Secondly, the HESA data on students' socio-economic background and their educational performance before university, which are strong predictors of student retention, is enormously more complete for young individuals than mature undergraduates.

The regression analysis results unveil that all ethnic minority groups are on average more likely to leave university because of academic failure (“involuntary” dropout) than White students. The picture is more alarming for Black African students, whose likelihood of failing academically (7.1%) is nearly two times higher than that of their White peers (3.8%). When looking at interaction effects, the ethnic gaps in academic failure are more extensive for men than women across most ethnic minority groups and are less pronounced in the Russell Group universities. Also, the ethnic disparities in involuntary dropout do not deviate markedly across different levels of socio-economic background.

On the contrary, White undergraduates have a greater chance of quitting voluntarily (for example, due to personal reasons or because they choose to go into employment) than all ethnic minorities. The corresponding difference ranges from 0.5 percentage point (when comparing White students with those from Mixed ethnic backgrounds) to 2.1 percentage points (White versus Black African students). In sum, the first-degree completion rates conditional on the total pool of observed characteristics are approximately one percentage point higher for White students (89.4%) relative to ethnic minorities, except for Indian (89.6%) and Chinese (89.5%).

Observing significant differences in dropout rates between ethnic groups suggests that policy actions regarding widening participation should further support degree accomplishment by ethnic minority students, instead of merely facilitating their access to universities. Hence, the UK universities would need to improve the degree completion rates of ethnic minorities while, in parallel, ensuring equitable access opportunities for all students. This becomes more challenging considering that there is evidence of a positive relationship between the institutions that attract a higher share of students from deprived and ethnic minority backgrounds and their average dropout rates (House of Commons, 2009; SMF, 2017).

Universities cannot wholly eliminate dropout, as there may always be a proportion of undergraduate entrants who soon find out that university is not for them or are not motivated enough to finish their studies (Tinto, 1982). In fact, some economists have inferred that eradicating dropout would not necessarily improve society, especially where the students' costs associated with attending university (financial or psychological) outweigh the potential benefits of receiving higher education (Hartog, Pfann and Ridder, 1989; Manski, 1989). However, the findings of this paper will inform policymakers about discrepancies in non-completion rates among certain ethnic groups with similar background characteristics and educational profiles. Hence, the present study aims to help decision-makers better understand the mechanisms that affect different outcomes and, consequently, underpin work to identify specific strategies that should be designed for each of those groups. On the universities side, examples of specific actions involve creating ethnically inclusive environments by enhancing the student support and counselling services, offering training sessions to develop the study skills of entrants with lower educational profiles, monitoring undergraduates' progress and adjusting the curriculum content, providing financial support for low-income individuals, and improving the information campaigns at schools through outreach programmes in order to align students' pre-entry expectations with their actual university experience.

The following section (2.2) reviews previous research investigating the drivers and inequalities in dropout decisions supported by the relevant theoretical framework. Section 2.3 provides descriptive analysis and presents the econometric strategy employed in this paper. In section 2.4, I discuss the results, and in section 2.5, I conclude and highlight the key policy implications of the current work.

2.2 Background and previous literature

2.2.1 Background

In general, the UK's degree completion rates are at good levels relative to international standards (Vossensteyn et al., 2015). To measure and compare universities' performance, the UK Government has developed some indicators that rely, among other characteristics, on dropout rates. For example, one of the six core metrics of the Teaching Excellence Framework (TEF), which was introduced in 2014 to reward and encourage high-quality teaching and partially defines the maximum amount of tuition fees that HE providers can charge in England, is based on universities' retention rates (BIS, 2016).

Dropout rates measure the proportion of individuals who leave their course for whatever reason before completing their studies. A well-known issue related to completion or dropout rates is that there are no universally accepted definitions of these terms. The central bodies, universities and the literature often use different ways to calculate student retention and withdrawal (McGivney, 1996; Smith and Naylor, 2002; Vossensteyn et al., 2015). Examining the probability of completion would ideally entail monitoring a chosen entrance cohort of undergraduates in a specified timeframe (for instance, from the commencement of their studies until a couple of years after their course's expected length). The major challenge with this "entry-cohort" method is that it necessitates use of several years of data to distinguish between completers and non-completers (especially for part-time students or courses that allow a high level of flexibility).

HESA's official continuation indicator is based on tracking the activity of UK-domiciled undergraduate entrants in the academic year following the entry year[2] (HEFCE, 2018; Advance HE, 2020). According to this measure, the proportion of Black students no longer in higher education stood at 10.3% in the academic year 2014/15 (down from 11.1% in 2010/11), which is 1.5 times bigger than that of White students (6.9% in 2014/15, up from 6.6% in 2010/11). In contrast, the Chinese and Indian minorities experienced the lowest dropout rates amongst all ethnic groups in 2014/15 (4.0% and 5.4%, respectively).

However, concentrating on early dropouts may produce an incomplete picture of the retention likelihood, given that students' withdrawal behaviour varies over time (Tinto, 1988; Smith and Naylor, 2002). For example, if, on average, White students remain in higher education longer than ethnic minorities, this method might overstate the overall ethnic gap in dropout rates. Moreover, some studies record as withdrawal events those students who transfer between institutions or switch programmes, while others separate out the individuals who leave higher education altogether (McGivney, 1996; Davies and Elias, 2003; Thomas and Hovdhaugen, 2014). In this paper, I adopt the "exit-cohort" approach suggested by some authors (e.g., Lee and Buckthorpe, 2008) to calculate the propensity to drop out or graduate by focusing on students who complete their studies or leave higher education in the same academic year.

From a theoretical viewpoint, a university degree may act as a signalling device of job seekers' quality in the labour market, especially in the early stages

[2] *HESA's indicator does not consider students who quit university very early (within 50 days of their studies' beginning). For part-time students, this indicator monitors their activity up to two years after they enrol at university.*

of their career. In his pioneering paper, Spence (1973) suggested that, in the absence of complete information about candidates' skills and capabilities, employers perceive educational credentials as a signal of the potential employees' productivity. The key condition required for Spence's signalling model to hold is that the signal cost is inversely associated with productivity (that is, the "good" workers' cost related to receiving additional education, which includes effort, time, emotional costs, and so forth, is lower than that of "bad-type" employees). If this hypothesis does not hold (that is, if everyone invests in education identically), the signal mentioned above cannot be exploited to distinguish candidates.

In economic literature, there has been a controversy between the traditional "human capital theory" (Becker, 1962) and the "screening hypothesis" of education (Arrow, 1973; Stiglitz, 1975). The former theory assumes that employees' productivity and socialisation increases through additional education, thus resulting in a higher market value of their labour. In contrast, education's filtering function proposes that earnings differences in the labour market do not necessarily mirror the productivity-improving effects of education, keeping all else equal. Instead, because of asymmetric information, employers consider education as a signal of pre-existing talents and skills of candidates, while prospective employees select their educational level to signal their ability to companies.

Because both theories have similar implications, there is limited empirical evidence separating the effects of human capital and screening on individuals' labour market prospects (Riley, 1979; Bedard, 2001). Some predictions of the screening theory, such as the "sheepskin effect", were not supported empirically at the time of its development (see a discussion in the paper of Layard and Psacharopoulos, 1974). Nonetheless, concerning the sheepskin-effect prediction, later evidence documented incremental returns to the acquirement of formal credentials, even when comparing workers of an equal number of education years (Hungerford and Solon, 1987; Jaeger and Page, 1996). Hence, from this perspective, the labour market position of university dropouts is more difficult compared to university graduates.

2.2.2 Previous research on student attrition

In his seminal work in the US, Tinto (1975) proposed that the likelihood of non-completion depends on a multidimensional process that defines the level of students' integration into university. Specifically, according to his path analysis model, which constituted the theoretical foundation for later empirical studies, there is a combination of personal, social, academic, and organisational forces that affect the probability of student retention in higher education (see also Tinto, 1993). Tinto's theoretical framework built on the notion of integration in Durkheim's (1897) sociological study on suicide and extended Spady's (1970) previous work. The subsequent literature has reviewed some key factors and their interrelationships that influence students' decision to leave university. These include individual and cultural characteristics, learning approaches, university practices, and policy-related determinants (Quinn, 2013).

There is a substantial amount of research that has quantified the effect of the chief characteristics determining the likelihood of non-completion in the UK higher education (e.g., Johnes, 1990; Yorke, 1998b; Bennett, 2003; Davies and Elias, 2003; Christie, Munro and Fisher, 2004; Arulampalam, Naylor and Smith, 2005; Yorke and Longden, 2008; Hovdhaugen and Aamodt, 2009; Rose-Adams, 2013; Collings, Swanson and Watkins, 2014; SMF, 2017; Webb and Cotton, 2018). More specifically, family background is a powerful driver of enrolment loss, as it is correlated with the students' financial capacity, their preparedness to succeed in higher education, and their post-university career aspirations (Johnes, 1990; Vignoles and Powdthavee, 2009). For example, Crawford (2014a) showed that students from the most deprived backgrounds have a 3.4 percentage points higher propensity to leave university within two years of commencement (2004-2006) compared to the top socio-economic quintile group, even after accounting for demographic characteristics, school attainment, type of university, and course fixed effects. Pre-university educational performance and the school type attended are also associated with completion rates (Fielding, Belfield and Thomas, 1998; Smith and Naylor, 2002; Arulampalam, Naylor and Smith, 2005).

Johnes and McNabb (2004) utilised information about 95,000 university leavers in 1993 ("exit cohort") to decipher the significance of peer effects and the university's quality in students' non-completion. Using multinomial logit models, the authors distinguished between academic failure and voluntary dropout. They found that prior educational attainment (measured by A-level scores) is positively associated with the probability of successful completion. Nevertheless, the authors showed that students of higher prior educational ability relative to their peers' average were more likely to drop out voluntarily than those with below-average previous educational attainment. Moreover, they indicated that the university's research and teaching quality is negatively linked with the likelihood of withdrawal. Interestingly, the authors found that UK-born and overseas students have a higher propensity to leave university than first-generation immigrants and "home" (UK-domiciled) students, respectively. Hence, they speculated (but did not show) that early-arriving immigrants, who are classified as home undergraduates, have the lowest chances of quitting university because of family pressure to succeed in higher education.

Drawing on an exit survey of 462 first-year undergraduates studying various subjects in 25 UK universities during the academic year 2005/06, Yorke and Longden (2008) used principal component analysis to spot the primary reasons for students' departure. The seven key influences detected in their study were the deficient learning experience; the wrong choice of the subject of study; the university's location and environment; problems with meeting the academic requirements; the lack of satisfaction with the resources provided; financial/employment difficulties; and issues related to students' social integration.

Most of the abovementioned studies neglect the role of ethnicity as a principal dimension of dropping out. While some of them control for broad ethnic groups in their analysis (by distinguishing, for example, between White and non-White students), they do not focus on the effect of ethnicity *per se* on the probability of academic dismissal or voluntary withdrawal. One possible explanation is related to the assumption that once accounting for specific critical factors, such as entry qualifications, socio-economic background, age, financial difficulties, peer effects, teaching quality, and social integration in the university environment, the impact of ethnicity on non-completion should be eliminated (Rodgers, 2013). However, in this paper, I show that significant ethnic differences in the likelihood of dropping out remain even after allowing for dissimilarities in many individual attributes and institutional characteristics.

Only a handful of papers have explicitly addressed the impact of ethnicity on student retention. In one of the earliest studies in the field, based on a sample of 858 HE students who enrolled in Bradford College during the period 1975-1982, Singh (1990) showed that the dropout rate was 35% for Black students and 28% for their White counterparts. Building on a survey undertaken in 2002/03 covering students of the second year and above, Connor et al. (2004) found that a higher proportion of Black undergraduates (36%-38%) considered quitting university early compared to the Chinese (24%), Indian (26%), Pakistani/Bangladeshi (27%) and White (28%) ethnic groups. However, this descriptive evidence base relied on a few respondents and did not control for socio-demographic and other student and university characteristics.

A more comprehensive report conducted by the National Audit Office (NAO) in 2007 used binary logistic regression models to estimate the students' probability of continuing to the second year of their studies, conditional on a wide range of characteristics (age, socio-economic background, ethnicity, gender, disability status, subject of study, type of institution, and prior educational attainment). By analysing the HESA data for the cohort of students who entered university in the 2004/05 academic year, the NAO report found that the odds ratios of staying in university were higher for most full-time ethnic minority groups relative to the White community. Specifically, the odds ratios favoured Indian and Chinese entrants (1.20), whereas students of unknown ethnicity (0.68) saw the lowest continuation propensity. On the contrary, the part-time entrants from ethnic minority backgrounds were less likely to continue to the second year of their course than their White peers, particularly those of Pakistani and Bangladeshi ethnic origins (0.81). Because the present study focuses on young undergraduates, the proportion of part-time students is relatively small (3%).

Vignoles and Powdthavee (2009) utilised linked National Pupil Database and HESA data for the academic years 2004/05-2005/06 to examine the likelihood of dropping out from the first year of the course for the pupils entering university at 18. Perhaps surprisingly, the authors found that all ethnic minority groups had a lower chance of withdrawal than their White counterparts, even after considering socio-economic and other personal attributes, prior educational achievement, subject of study, and university fixed effects. Specifically, the corresponding ethnic gap ranged from 2.1 percentage points in favour of Pakistani, Bangladeshi, and Black Caribbean students to 3.2 percentage points in favour of Chinese entrants. This picture might be attributable to the fact that the authors could not differentiate between voluntary withdrawal and involuntary dropout because of data limitations. Hence, these types of withdrawal likely cancelled each other out, leading to a small net effect in favour of ethnic minorities. The present paper demonstrates that students' dropout behaviour differs markedly between White and ethnic minority groups depending on whether it is triggered by academic failure or other factors (such as financial or personal reasons).

2.3 Data and Methodology

I use pooled HESA data[3] for the academic years 2010/11–2014/15. The HESA datasets cover all students in the UK universities and contain extensive personal-level information about socio-demographic characteristics, parental education, prior attainment, subject and mode of study, expected length of programme, source of tuition fees, term-time accommodation, region and type of university, and pre-entry characteristics (e.g., the type of school attended). I construct some university-level variables (such as the staff/student ratio and the non-White/White staff ratio) using information from the publicly available HESA student and staff records and merge them with my main HESA datasets. I also collected the publicly available Teaching Excellence Framework outcomes from the Office for Students, as updated in June 2020 (OfS, 2020).

Table 2.A1 in the Appendix presents a comprehensive description of all variables incorporated in this analysis to estimate the probability of first-degree completion, involuntary attrition, and voluntary dropout. The selection of independent variables is based on the previous studies discussed in section 2.2. As I describe in section 2.3.3, I consider three outcomes: degree completion, compulsory withdrawal because of academic failure, and dropout caused by other reasons. Following Johnes and McNabb (2004), I describe academic dismissal as “involuntary attrition”, whereas “voluntary dropout” groups together non-completers who left university due to personal or financial reasons, departure to employment, exclusion, writing off after a period of inactivity, health reasons, or death.

2.3.1 Sample selection

The sample comprises 1.4 million first-degree students who were UK-domiciled before commencing their course. It includes only students who ended their course because they successfully completed their studies in any of the academic years 2010/11-2014/15 or because they left their course for whatever reason (e.g., academic failure or other reasons). The ethnicity data is available

[3] *“HESA Student Record 2014/15; HESA Student Record 2013/14; HESA Student Record 2012/13; HESA Student Record 2011/12; HESA Student Record 2010/11. HESA declaration: Copyright Higher Education Statistics Agency Limited. Neither the Higher Education Statistics Agency Limited nor HESA Services Limited can accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by HESA Services.”*

only for those undergraduates who lived in the UK before beginning their studies. Hence, the fact that the sample consists of students who were schooled in the UK should minimise the effect of any unobserved disadvantages in commanding the English language between ethnic groups or differences in curricula across countries.

To improve the homogeneity of my sample, I have dropped all mature students (aged over 21 on entry), distance learners, “other undergraduates” (such as individuals aiming for “Higher National Diploma”, “Certificate of Higher Education”, and “Higher National Certificate”), and postgraduate students. The information about some crucial factors that determine academic failure or the decision to quit higher education (such as previous performance and socio-economic classification) is considerably narrowed for mature students in the HESA datasets. For example, the missing values for the *tariff score* variable (which measures the pre-university educational attainment) cover 84% of mature students, while the corresponding figure relating to their socio-economic background is 52%. Moreover, the HESA data lacks information about some factors that probably influence mature students’ academic attainment and motivation to accomplish their studies to a larger degree than that of young students (e.g., family formation, income, aspirations, and experiences) and which likely differ amongst ethnic groups. Consequently, including mature students in the sample without controlling for these characteristics in the regression models would exacerbate the omitted variable bias in the estimated effect of ethnicity on the likelihood of dropping out of higher education.

The resultant sample does not contain students transferred to another institution. The principal reason for this is that students who choose to transfer to another provider are not considered university dropouts, as they remain in the higher education sector. Therefore, this group of students is outside the scope of the present analysis[4]. The second reason is associated with technical issues in the HESA datasets. Specifically, for each academic year, students have a unique anonymous identifier. However, if students change course or provider, they will

[4] *When including students transferred to another institution in the sample, their share covers 0.75% of the total sample. The proportion of individuals transferring to another university does not differ much between ethnic groups (White: 0.72%, Black: 0.89%, Chinese: 0.66%, Other Asian: 0.90%, Mixed: 0.88%, Other ethnic group: 1.05%, Unknown ethnicity: 0.67%). Therefore, excluding those individuals from the present analysis should not bias the results regarding ethnicity’s effect on the likelihood of retention/withdrawal.*

also likely change identifiers[5]. Therefore, embodying transfers to other universities in the sample would likely lead to double counting of some students. Moreover, I cannot track students from one year to the next using this anonymous identifier. Hence, as mentioned earlier, the current work relies on individuals who gained their qualification or left university altogether in the same academic year ("exit cohorts").

The exit-cohort approach has the advantage that it allows the students' withdrawal behaviour to vary over time. For example, some ethnic minorities might encounter insurmountable hindrances to integrate socially and bond with their university at the early stages of their academic life, or they might lack specific skills on entry required to succeed compared to other ethnic groups, resulting in diverse effects of ethnicity on their performance. Indeed, among the students in the sample who failed academically, 35% of Chinese did so in the first year of their studies, which is considerably lower than the respective figure of Bangladeshi (51%), Pakistani (46%), Black (45%), and White (44%) students.

On the other hand, the exit-cohort method implies that many students have started university in different years. For instance, graduates of part-time courses should have enrolled at university earlier than full-time dropouts. As a result, if the characteristics and the proportion of students who successfully progress in their course fluctuate across years, then the exit-cohort method used here might generate different dropout rates than the approach focusing on students who joined the course in the same academic year ("entry cohorts"). In the present study, I attempt to reduce this sampling bias originating from mixed entry cohorts by conditioning on a rich set of personal traits, previous educational attainment, and students' year on course.

2.3.2 Descriptive evidence

Table 2.A2 in the Appendix summarises the average socio-demographic and academic characteristics of White and ethnic minority (non-White) students

[5] *Because of the high level of granularity in the requested variables (e.g., detailed ethnicity classification, age, gender, socio-economic classification, domicile, institution attended), HESA's data protection team did not provide me with the 13-digit unique student identifier, which would allow tracking students across all years of their university life. Instead, the anonymous identifier used here only permits distinguishing undergraduates within the same academic year and institution. Hence, although I can tell whether a student transferred to another university, I cannot track their progression once they change provider. Similarly, with the present datasets, it is not possible to differentiate between students who drop out of higher education altogether and those who quit university in a given academic year but re-apply through the admissions service for enrolling at a new course/university the years thereafter.*

included in the sample. The raw (unconditional) completion rate of non-White students in the academic years 2010/11–2014/15 is 3.7 percentage points lower than that of their White peers, standing at 86.1%, on average. This ethnic gap in the graduation likelihood is predominantly driven by academic performance (see also Figure 2.1). Specifically, 7.9% of ethnic minorities fail assessments at any stage of their studies and are not allowed to progress further, which is more than double the respective rate for White students (3.5%). Conversely, White students are slightly more likely to drop out voluntarily (6.7%) than non-White undergraduates (6.0%). In particular, the White group sees a higher proportion of withdrawal due to personal reasons or because students depart their course to go into employment.

On average, ethnic minorities are more likely to study STEM (38.6%) and LEM (26.8%) subjects compared to their White fellows (35.6% and 17.8%, respectively). Nevertheless, the former are not equally represented in the prestigious Russell Group universities (21.3% versus 27.6%). This is also reflected by the fact that ethnic minorities attend universities with a considerably lower average tariff score at entry (which is a measure of selectivity) than White students (329 versus 348 tariff points). Moreover, non-White students primarily choose universities in London (33.0% compared to only 7.7% of White students). This is directly correlated with their “average distance travelled” (71 km versus 109 km) and the fact that 42% of non-White students live in their family home during term time, considering that a vast proportion of ethnic minority communities reside in London.

The parents/guardians of ethnic minority students are less likely to hold a higher education qualification, and only 33.9% of them have managerial or professional jobs, compared to 48.1% of White undergraduates’ parents. In addition, a smaller percentage of non-White students (8.6%) attended private schools before enrolling at university relative to their White fellows (11.3%). The latter figures taken together support the argument that, on average, ethnic minority students come from lower social classes than their White counterparts. In the multinomial regression analysis (section 2.4), I control for these background characteristics to quantify the ethnic discrepancies in the probability of degree completion, compulsory withdrawal, and voluntary dropout.

Table 2.1 shows the detailed distribution of non-continuation reasons for each of the eleven ethnic groups. Of the total 1,389,920 young, UK-domiciled, first-degree students with a known reason for ending their course, 79.6% identified as White, and one fifth were from ethnic minority backgrounds. 1,238,100 students (89.1% of the total sample) successfully completed their studies, 60,760 students (4.4%) failed their assessments, and 91,060 undergraduates (6.6%) dropped out for any other reason (voluntary leavers). On average, Chinese (92.8%), Indian (90.8%), and White (89.8%) students show the highest completion rates among all ethnic groups. The students from Pakistani, Bangladeshi, other Asian, and Mixed ethnic backgrounds are more likely to complete their course than Black minorities, whose completion rates stand at relatively low levels (79%-82%).

Black students are three to four times more likely to fail assessments (10%-12%) than their White counterparts (3.5%). Similarly, a higher share of Black students leave their course because of financial reasons, infringement of university's behavioural/financial regulations (exclusion), or inactivity (written off) compared to all other ethnic groups. Research has shown that financial hardship affects the decision to leave university (Bennett, 2003; Connor et al., 2004), while students from ethnic minorities and poorer backgrounds are more debt-averse than others (Callender, 2003; Callender and Jackson, 2005). In this case, a policy solution would be to improve further the awareness of prospective entrants about financing arrangements in higher education and the associated costs of attending a degree course.

On the contrary, a higher percentage of White students than those from ethnic minorities (except for students from Mixed ethnic backgrounds) drop out for personal reasons in the period considered. For example, an average White student is twice more likely (3.3%) to leave voluntarily due to personal reasons than a Chinese undergraduate (1.6%). Moreover, as mentioned, White people have a stronger possibility of leaving university willingly to enter employment (0.4%) than ethnic minorities. This might suggest that the average opportunity cost of higher education is greater for White individuals, particularly those living in regions with low unemployment rates (for a discussion, see Giani, Attewell and Walling, 2020). It could also imply that a higher proportion of White students expect lower labour market returns to their university degree than their ethnic minority peers and, therefore, have reduced incentives to finish their studies.

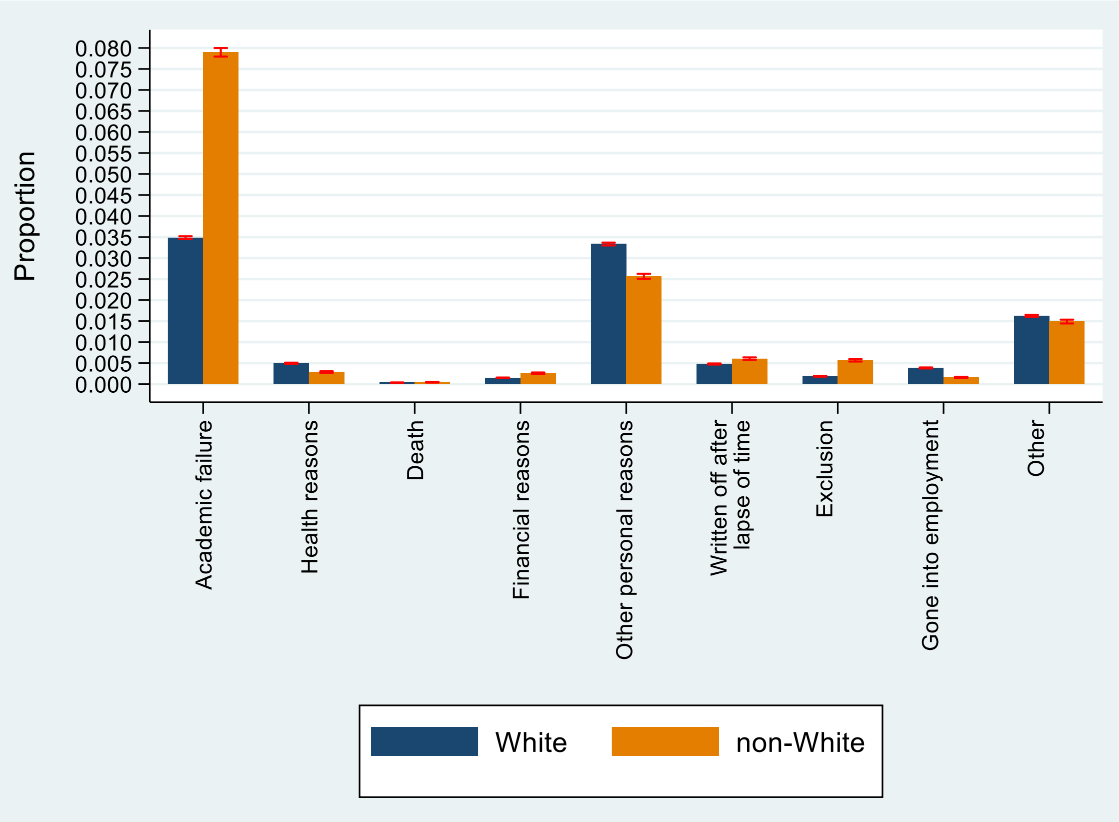
Tables 2.2-2.4 present the cross-tabulations of first-degree students who failed academically (involuntary attrition) by ethnicity and institution type, subject of study, and socio-economic background. Tables 2.A3-2.A5 in the Appendix report the corresponding figures for voluntary leave students (in a grouped form) who quit higher education due to any other reason. On average, students attending a Russell Group university are less likely to fail academically (1%) than their fellows studying at other pre-1992 (4%), post-1992 (6%), and specialist (3%) institutions (see Table 2.2). Across all types of universities, White students exhibit the lowest failure rates (ranging from 1% to 5%), followed by the Chinese ethnic group (2% to 6%). In contrast, Black, Pakistani, and Bangladeshi undergraduates experience the highest probability of academic failure, which is more than twice as large as that of their White peers.

Similarly, the proportion of White students who do not meet the academic standards and withdraw is smaller than that of ethnic minorities across most subject areas of study (Table 2.3). More specifically, undergraduates in Computer science (11%), Engineering and Technology (8%), and Architecture (7%) are the most likely to fail academically. Again, within those subject areas, the probability of failing is more remarkable for Black, Pakistani, and Bangladeshi students (ranging from 14% to 22%) than their White counterparts (5%-9%).

Moreover, there is an evident relationship between students' socio-economic classification (proxied by parental occupation) and their involuntary non-completion rates (Table 2.4). For instance, the average propensity for failure is 3% for undergraduates whose parents hold a higher managerial/professional position, while the respective probability for students whose parents work in routine occupations (7%) or are long-term unemployed (10%) is significantly greater. In section 2.4, I explore interaction effects to investigate whether there are statistically significant ethnic differences in the likelihood of failing academically across various types of institutions and segments of the socio-economic distribution, conditional on the total pool of observed characteristics.

As alluded to in section 2.2, earlier studies have found that prior educational aptitude is positively correlated with graduation probability. The UCAS tariff score is an appropriate measure of the previous achievement, as it captures the students' qualifications obtained after the age of 16 and the associated grades in a single indicator. As Figure 2.2 depicts, prior attainment of Black, Bangladeshi, and Pakistani students falls significantly below that of White and Chinese students, on average. This suggests that the performance gaps of specific ethnic minorities may develop at earlier stages of their educational life, thus undermining their propensity to succeed at university. In the following regression analysis, I interact ethnicity with UCAS tariff scores to account for the fact that some types of students are more likely to fail academically once they enter university, depending on the level of their prior educational ability.

Figure 2.1. Proportion of dropouts by reason for ending course: White versus non-White students



Note: The error bars represent the 95% confidence interval of the corresponding proportion for each ending reason.
The total sample of White and non-White UK-domiciled young undergraduates is 1,106,695 and 270,645, respectively.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.1. Distribution (%) of non-continuation reasons by ethnic group

Ethnic group	Reason for ending course										
	Successful completion	**Academic failure**	**Health reasons**	**Death**	**Financial reasons**	**Other personal reasons**	**Written off after lapse of time**	**Exclusion**	**Gone into employment**	**Other**	**n**	**%**
White	89.8	3.5	0.5	0.04	0.2	3.3	0.5	0.2	0.4	1.6	1,106,695	79.6%
Black Caribbean	82.1	10.2	0.4	*	0.5	3.2	0.9	0.7	0.2	1.8	17,145	1.2%
Black African	81.5	12.0	0.2	*	0.5	2.3	0.9	1.1	0.1	1.5	42,495	3.1%
Other Black	79.2	12.3	0.4	*	0.4	2.9	1.4	0.9	0.1	*	3,100	0.2%
Indian	90.8	5.2	*	*	*	1.8	0.4	0.2	0.2	1.1	57,610	4.1%
Pakistani	85.0	8.5	0.3	*	0.3	2.9	0.5	0.6	0.2	1.7	39,885	2.9%
Bangladeshi	84.6	9.1	0.2	*	0.2	3.0	0.6	0.8	0.2	1.5	15,645	1.1%
Chinese	92.8	3.7	0.2	*	0.1	1.6	0.3	0.2	0.1	*	13,895	1.0%
Other Asian	85.8	8.7	0.2	*	*	2.2	0.6	0.5	0.2	1.6	20,820	1.5%
Mixed	86.5	6.1	0.5	0.06	0.2	3.5	0.7	0.5	0.2	1.8	46,500	3.3%
Other ethnic group	84.9	9.0	0.4	*	0.2	2.6	0.6	0.7	0.2	1.6	13,550	1.0%
Unknown	87.3	6.5	0.5	*	0.2	3.0	0.6	0.4	0.2	1.3	12,580	0.9%
Total	**89.1**	**4.4**	**0.5**	**0.04**	**0.2**	**3.2**	**0.5**	**0.3**	**0.3**	**1.6**	**1,389,920**	**100%**

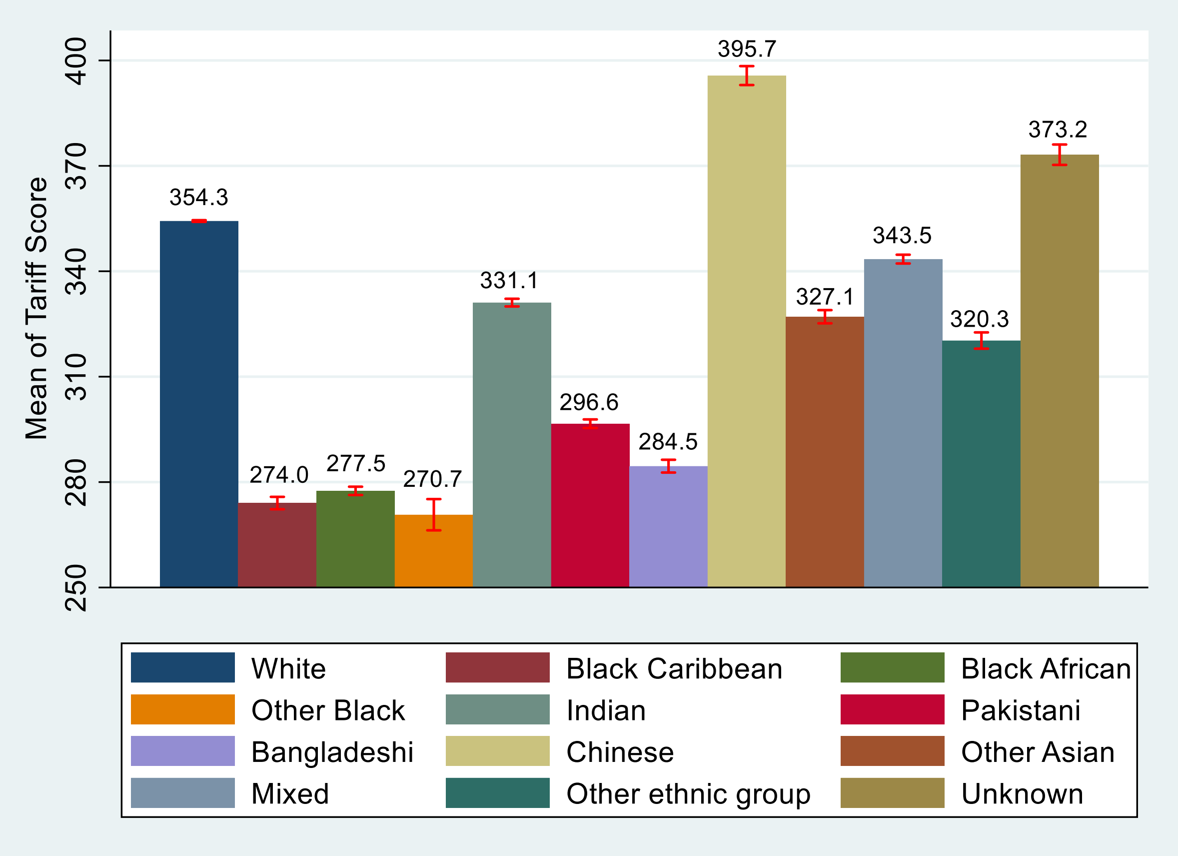
Note: Percentages in each row sum up to 100%.
** denotes cells with fewer than 23 students. The total number of students (n) for each ethnic group is rounded to the nearest multiple of 5, in line with the data provider's disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

Table 2.2. Proportion of students who failed academically by ethnic group and type of university

Ethnic group	University type				
	Russell Group	Other Pre-1992	Post-1992	Specialist	**Total**
White	0.01	0.03	0.05	0.03	**0.03**
Black Caribbean	0.03	0.07	0.12	0.08	**0.10**
Black African	0.04	0.08	0.15	0.14	**0.12**
Other Black	*	0.08	0.15	*	**0.12**
Indian	0.02	0.05	0.07	*	**0.05**
Pakistani	0.04	0.07	0.10	*	**0.08**
Bangladeshi	0.03	0.09	0.11	*	**0.09**
Chinese	0.02	0.04	0.06	*	**0.04**
Other Asian	0.03	0.09	0.11	*	**0.09**
Mixed	0.01	0.04	0.09	0.06	**0.06**
Other ethnic group	0.03	0.07	0.12	*	**0.09**
Unknown	0.01	0.05	0.12	*	**0.07**
Total	**0.01**	**0.04**	**0.06**	**0.03**	**0.04**

*Note: * denotes cells with fewer than 23 students.*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.2. Average tariff score by ethnic group



Note: The error bars represent the 95% confidence interval of the mean tariff score of each ethnic group.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.3. Proportion of students who failed academically by ethnic group and subject of study

Subject of Study	Ethnic group													
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Medicine & dentistry	0.01	*	*	*	0.03	0.04	*	*	*	0.02	*	*	**0.02**	21,100	1.5%
Subjects allied to medicine	0.03	0.08	0.10	*	0.04	0.06	0.06	*	0.05	0.05	0.06	0.05	**0.03**	86,485	6.2%
Biological sciences	0.04	0.10	0.13	0.14	0.05	0.08	0.07	*	0.08	0.07	0.11	0.07	**0.05**	148,050	10.7%
Veterinary science	*	*	*	*	*	*	*	*	*	*	*	*	*	2,340	0.2%
Agriculture & related subjects	0.04	*	*	*	*	*	*	*	*	*	*	*	**0.04**	9,025	0.6%
Physical sciences	0.03	0.12	0.15	*	0.05	0.10	*	*	*	0.05	0.06	0.05	**0.04**	61,140	4.4%
Mathematical sciences	0.03	*	*	*	0.05	0.09	0.08	0.04	*	0.06	0.06	*	**0.04**	24,885	1.8%
Computer science	0.09	0.19	0.22	0.19	0.11	0.15	0.21	0.08	0.16	0.16	0.17	0.16	**0.11**	53,945	3.9%
Engineering & technology	0.06	0.22	0.18	0.19	0.09	0.14	0.15	0.05	0.14	0.10	0.15	0.11	**0.08**	66,940	4.8%
Architecture, building & planning	0.05	0.15	0.18	*	0.10	0.15	0.16	*	0.15	0.10	0.17	0.10	**0.07**	28,590	2.1%
Social studies	0.03	0.08	0.10	*	0.03	0.07	0.07	*	0.06	0.05	0.07	0.06	**0.04**	95,345	6.9%
Law	0.03	0.07	0.07	*	0.04	0.08	0.08	*	0.07	0.06	0.08	0.07	**0.04**	52,625	3.8%
Business & administrative studies	0.04	0.11	0.12	0.13	0.06	0.10	0.11	0.05	0.10	0.09	0.10	0.10	**0.06**	124,465	9.0%
Mass communications & documentation	0.04	0.08	0.12	*	0.03	0.09	*	*	*	0.07	0.06	*	**0.04**	35,630	2.6%
Languages	0.02	0.06	0.09	*	0.03	*	*	*	*	*	0.06	*	**0.02**	58,300	4.2%
Historical & philosophical studies	0.02	*	0.08	*	*	0.06	*	*	*	*	0.03	0.04	**0.02**	54,810	3.9%
Creative arts & design	0.03	0.08	0.12	0.10	0.05	0.08	0.06	0.04	0.07	0.05	0.06	0.07	**0.03**	150,645	10.8%
Education	0.02	0.11	0.14	*	0.05	0.05	0.05	*	0.09	*	0.11	*	**0.03**	50,080	3.6%
Combined	0.03	0.10	0.11	0.13	0.04	0.08	0.08	0.03	0.07	0.05	0.08	0.04	**0.04**	265,520	19.1%
Total	**0.03**	**0.10**	**0.12**	**0.12**	**0.05**	**0.08**	**0.09**	**0.04**	**0.09**	**0.06**	**0.09**	**0.07**	**0.04**	**1,389,920**	**100%**

*Note: * denotes cells with fewer than 23 students. The total number of students (n) for each subject of study is rounded to the nearest multiple of 5, in line with the data provider's disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

Table 2.4. Proportion of students who failed academically by ethnic group and socio-economic background (parental occupation)

Socio-economic classification	Ethnic group														
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Higher managerial/professional	0.02	0.09	0.08	0.09	0.03	0.06	0.07	0.04	0.04	0.04	0.04	0.03	**0.03**	282,850	20.4%
Lower managerial/professional	0.03	0.10	0.11	0.11	0.05	0.08	0.08	0.04	0.07	0.05	0.08	0.05	**0.04**	346,510	24.9%
Intermediate	0.03	0.09	0.09	0.09	0.05	0.08	0.07	0.03	0.07	0.06	0.07	0.05	**0.04**	146,845	10.6%
Small employers/Own account workers	0.04	0.10	0.16	*	0.06	0.09	0.10	0.04	0.12	0.06	0.10	0.06	**0.05**	89,980	6.5%
Technical/lower supervisory	0.04	0.10	0.14	*	0.05	0.09	*	*	0.10	0.08	0.11	*	**0.04**	56,450	4.1%
Semi-routine	0.05	0.12	0.13	0.14	0.06	0.09	0.09	0.04	0.10	0.08	0.10	0.07	**0.06**	136,760	9.8%
Routine	0.06	0.11	0.16	0.20	0.06	0.08	0.09	*	0.11	0.09	0.11	*	**0.07**	64,270	4.6%
Long-term unemployed/Never worked	0.10	*	0.14	*	*	0.11	*	*	*	*	*	*	**0.10**	2,710	0.2%
Unknown	0.04	0.11	0.13	0.14	0.06	0.09	0.09	0.04	0.10	0.07	0.11	0.09	**0.06**	263,545	19.0%
Total	**0.03**	**0.10**	**0.12**	**0.12**	**0.05**	**0.08**	**0.09**	**0.04**	**0.09**	**0.06**	**0.09**	**0.07**	0.04	**1,389,920**	**100%**

*Note: * denotes cells with fewer than 23 students. The total number of students (n) for each level of socio-economic classification is rounded to the nearest multiple of 5, in line with the data provider's disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

2.3.3 Methodology

2.3.3.1 Multinomial logistic regression specification

I analyse a qualitative event with three unordered outcome categories. Specifically, the dependent variable (*Y*), which captures the reason for ending the first-degree course, is nominal and generates three possible discrete outcomes: successful degree completion, involuntary attrition, and voluntary dropout. I follow Johnes and McNabb's (2004) approach to consider as voluntary dropouts students who quit university due to personal or financial reasons, departure to employment, exclusion, writing off after a period of inactivity, health issues, or even death[6]. Involuntary attrition refers to compulsory withdrawal and covers the students who left higher education due to academic failure.

Multinomial logistic regression (MLR) is the most common method used in the literature to synchronously estimate binary models for all possible comparisons between the response variable's groups. Formally, the predicted probabilities (π_{im}) of the MLR model (McFadden, 1974; Long and Freese, 2014) are expressed as:

$$\pi_{im} = Prob(Y_i = m|X_i = x_i) = \frac{\exp(\mathbf{x'}_i\boldsymbol{\beta}_{m|\mathrm{b}})}{\sum_{j=1}^{J}\exp(\mathbf{x'}_i\boldsymbol{\beta}_{j|\mathrm{b}})} \quad (1)$$

where m =1, 2, …, J are the dependent variable's outcomes of student i; b represents the base category (comparison group); $\mathbf{x}_i$ is the vector of explanatory variables, and $\boldsymbol{\beta}$ represents the unknown parameters. In the present analysis, because J=3, only two (J-1) binary logits are estimated independently, as the sum of the probabilities of all outcomes should be one. Specifically, by setting b=1 ("successful degree completion") as the reference category, the MLR models produce the estimates $\hat{\beta}_{2|1}$ and $\hat{\beta}_{3|1}$ (where $\beta_{1|1} = 0$, given that the logarithm of the odds of an outcome relative to itself is zero). Nevertheless, the estimated propensities will be the same irrespective of the choice of the reference category.

[6] *As alluded to in the previous subsection (2.3.2), death is a rare occurrence in the data. As I will show in the robustness checks section (2.4.3), removing students who did not complete their studies due to health reasons/death from the sample (as they are not "voluntary" dropouts in the precise sense of the term) makes very little difference to the effect of ethnicity on the likelihood of dropping out.*

Apart from the key variable of interest (ethnicity), the vector $\mathbf{x}_i$ comprehends all the variables described in the previous section and the interaction terms discussed in the following subsection. Specifically, it includes socio-demographic traits (gender, age, disability status, socio-economic background, and parental education); institutional and study characteristics (type and region of the university, subject of study, mode of study, programme's length, student's year on the course); pre-entry factors (tariff score, type of school, distance travelled from student's home before entry to the institution); other individual-level characteristics (term-time accommodation, home fees eligibility, source of tuition fees); peer effects (proportion of non-White peers, relative tariff score); university quality measures (TEF outcome, staff-student ratio, non-White/White staff ratio, university's income per student); and academic year fixed effects.

2.3.3.2 Interaction effects

Logit models are interactive in nature because of the sigmoid logistic function. This means that the effect regressors have on the dependent variable is not identical across different places of the function's curve (even when the models lack product terms), given that the estimated probabilities are restricted between zero and one. However, failing to incorporate a significant interaction term in the model would bias the results, as this would not permit the slopes of this function to fluctuate across different levels of the independent variables, thus compromising the conclusions regarding the real interactive relationship (Rainey, 2016).

To explore whether the impact of ethnicity on the likelihood of dropping out is heterogeneous depending on students' characteristics, I include several interaction terms in the MLR models. In section 2.4, I focus on the interaction effects relating to academic failure, as the ethnic differences are more pronounced within this dependent variable's category. In this respect, the vector $\mathbf{x}$ in specification (1) also encompasses the following product terms of interest: "*Ethnicity*Gender*", "*Ethnicity*University type*", "*Ethnicity*Socio-economic classification*", and "*Ethnicity*Tariff Score*". For instance, the interaction effect between ethnicity and gender denotes whether the impact of ethnicity on the propensity to fail academically differs between men and women. The MLR models also contain the quadratic term of the variable "*relative tariff score*", which captures the students' relative educational ability as a fraction of their peers' average ability. In doing so, I consider the potential differential impact of academic peer effects on involuntary attrition at various levels of the previous educational performance. For example, one would anticipate that the likelihood of dropping out is smaller for undergraduates whose educational aptitude is close to that of their fellows than other students. I approximate the peers' academic ability by computing the average tariff score of students within the same university, course, and academic year.

Ai and Norton (2003) emphasised that, in non-linear models, the interaction effect cannot be assessed by merely looking at the statistical significance, sign, and coefficient size of the corresponding interaction terms produced by the original regression (see also Buis, 2010). Instead, to obtain consistent estimators for the interaction effects in non-linear specifications, one should compute cross derivatives (or cross differences, with factor regressors) of *Y*'s conditional expected value. Therefore, in the present work, I estimate the marginal effects to test and quantify the interaction effects mentioned above.

2.3.3.3 Marginal effects

Marginal effects (MEs) synopsise the impact of ethnicity on the dependent variable in a single measure, even when the models include product terms (Long and Freese, 2014). They are calculated using the predicted probabilities derived from the original multinomial logistic regression (post-model estimation). MEs are an easily interpretable metric of each ethnic minority's effect on the likelihood of graduation, academic failure, and voluntary dropout.

This measure also addresses the well-documented identification (scaling) issue in models with latent dependent variables, which renders the coefficients equality tests invalid (Lee, 1982; Mood, 2009). Specifically, the fact that the unit of the *Y* variable does not have a scale (such as pounds, kilograms, or test scores) implies that the unit can be fixed only by keeping the residual variance constant. Hence, if the error term variance changes (by adding, for example, a new variable in the model), then this would also affect the scale of *Y*. As a result, the impact of the other independent variables on *Y* would probably change, even if the new independent variable added in the MLR model is uncorrelated with the other observed regressors, thus rendering the comparison of coefficients between models problematic. In the current setting, this problem prevents comparing the MLR coefficients across groups of students. For instance, if the unobserved characteristics influence the probability of dropping out disproportionately between men and women, then the scale of the dependent variable would differ across genders. Comparing predicted probabilities (instead of the raw MLR coefficients) provides a remedy for this problem (Long and Mustillo, 2018).

In the current context, average marginal effects (AMEs) are particularly useful for estimating and reporting the interaction effects between ethnicity and other explanatory variables (e.g., gender, university type, and social class). For each ethnic minority group ($ethnic = k$), the AMEs represent the average difference in the predicted probabilities of outcome m for all students in the sample (π_{im}) between the ethnic group k and the White reference group, conditional on the observed values in the data of the rest explanatory variables (equation 2).

$$AME_{m,ethnic_k} = \frac{1}{N}\sum_{i=1}^{N}(\pi_{im}\,(ethnic = k, \mathbf{x} = \mathbf{x}_i) - \pi_{im}\,(ethnic = White, \mathbf{x} = \mathbf{x}_i)) \quad (2)$$

To examine whether the AMEs for each ethnic group are equal across selected levels of other independent variables (that is, to test the interactive effects), I use the “second differences” approach (Mize, 2019). For instance, let $\hat{\Delta}_{ethnic_k|women}$ and $\hat{\Delta}_{ethnic_k|men}$ be the AMEs of the ethnic minority group k concerning academic failure for women and men, respectively. A second-difference test shows whether the difference in the likelihood of academic failure between the ethnic group k and the White reference category differs statistically significantly between women and men. The denominator in equation (3) represents the estimated standard errors and covariance between the two marginal effects. Statistical software packages generally use the *delta method* (Agresti, 2013; Dowd, Greene and Norton, 2014) to compute the variances of partial effects, while a Wald test examines whether these marginal effects are equal.

$$z = \frac{\hat{\Delta}_{ethnic_k|women} - \hat{\Delta}_{ethnic_k|men}}{\sqrt{\hat{\sigma}^2_{ethnic_k|women} + \hat{\sigma}^2_{ethnic_k|men} - 2\hat{\sigma}_{ethnic_k|women\,,\,ethnic_k|men}}} \quad (3)$$

In section 2.4, I also report the average adjusted predictions (AAPs) for each ethnic group, which represent the respective conditional probability of successful completion, voluntary dropout, and involuntary withdrawal, keeping the rest regressors' values as is. AAPs are substantively useful to predict the average likelihood of these three outcomes for each ethnic group after controlling for the complete set of observable characteristics.

2.3.3.4 Testing the multinomial logistic regression assumptions

It is usually an empirical matter to decide if a subgroup of the dependent variable's alternatives should be considered a single outcome in MLR models. I run a Wald[7] test (Long and Freese, 2014) to formally assess whether I should combine any alternatives of *Y* into one category. Specifically, the null hypothesis of this test is that none of the observed regressors x substantially influences the odds of the outcome m versus the outcome n, suggesting that m and n are indistinguishable with regard to the covariates:

$$H_0: \beta_{1,m|n} = \cdots \beta_{x,m|n} = 0$$

If the null hypothesis is not rejected, then combining the corresponding pair of states into one category (for example, merging the "involuntary attrition" and "voluntary dropout" alternatives) would provide more efficient estimates. However, as Table 2.A6 in the Appendix shows, the chi-square statistic is very large for all possible combinations, thus rejecting the null hypothesis that any two alternatives should be pooled.

A critical assumption pertaining to the MLR models is the independence of irrelevant alternatives (IIA). This assumption states that the odds of an individual's preference between two different outcomes of the dependent variable (m and n) do not depend on the existence of other alternatives. In other words, including a new alternative outcome in the dependent variable should not alter the relative propensities of the previous choices. For example, the odds of failing academically versus leaving university voluntarily due to personal reasons should remain the same after including a different alternative (e.g., leaving university

[7] *Alternatively, one could use the likelihood-ratio (LR) test (Cramer and Ridder, 1991). However, apart from the fact that the LR test is computationally costly (especially for extensive datasets and models with many explanatory variables, as in the present case), it cannot be utilised when the MLR models use robust standard errors.*

because of financial reasons). Similarly to the classical linear models, this assumption requires that the omitted characteristics from the MLR specification are independent random variables (that is, the unobserved error terms are uncorrelated across different alternatives). In practice, the IIA property is violated when the alternatives are close substitutes, resulting in inconsistent estimators of the true population parameters[8].

There are a few tests of IIA comparing the parameter estimates from the complete regression to those from the refitted model after removing one of the outcomes of the dependent variable (Hausman and McFadden, 1984; Small and Hsiao, 1985). However, some authors strongly oppose using these tests (Fry and Harris, 1998; Long and Freese, 2014), contending that they are unsuitable for evaluating violation of the IIA assumption. The reason is that these tests often produce contradictory results, even when working with the same data and models, while simulations have shown that they have poor size properties (Cheng and Long, 2007).

Indeed, as Table 2.A7 in the Appendix indicates, two different IIA tests lead to conflicting conclusions. The Hausman-McFadden (HM) test seems not to reject the null hypothesis that the odds of outcome m versus the outcome n are independent of other alternatives, thus satisfying the IIA property. Specifically, the statistic used for the HM test is:

$$\mathrm{HM} = (\hat{\beta}_R - \hat{\beta}_U^*)' \, [\widehat{Var}(\hat{\beta}_R) - \widehat{Var}(\hat{\beta}_U^*)]^{-1} \, (\hat{\beta}_R - \hat{\beta}_U^*) \tag{4}$$

where $\hat{\beta}_R$ represents the coefficients in the reduced model after removing one or more states of the dependent variable; $\hat{\beta}_U$ are the coefficients in the unrestricted model; and $\hat{\beta}_U^*$ is a subgroup of $\hat{\beta}_U$ after dropping the coefficients that do not fit

[8] *A typical example presented in the literature, which renders this property unrealistic, refers to the transportation modes used to commute to work (McFadden, 1974). Imagine that people's choices are going to work by car or on a blue bus. Supposing that these two means of transport have equal probabilities (50% each), the corresponding odds ratio will be one. Now assume that the alternative options increase and a red bus becomes available. It would be realistic to anticipate that the likelihood of riding a red bus would be the same as that of getting a blue bus, as these choices are perfect substitutes. In this case, the IIA would hold only if the initial odds ratio remains unchanged (that is, only if the likelihood of taking a car is 33%, a blue bus 33% and a red bus 33%). However, this is highly unlikely to be the case in reality because the probability of taking a car would most probably remain 50% even after introducing the new alternative, thus violating the IIA assumption (50%/25%=2≠1).*

in the reduced model. In contrast, the Small-Hsiao test provides evidence that the IIA assumption is violated (Table 2.A7 in the Appendix).

McFadden (1974) and Amemiya (1981) argued that the MLR models perform well when the dependent variable's states are distinct, recommending that researchers should not choose sets of similar alternatives. In the present study, I assume that successful degree completion, voluntary dropout, and involuntary attrition are not substitutes for one another. The mechanisms and earlier processes associated with each of those outcomes should differ in character (Larsen, Sommersel and Larsen, 2013). For instance, as I will show in section 2.4, socio-economic background and previous attainment are stronger predictors of academic failure than voluntary withdrawal. Also, unlike compulsory dismissal due to academic failure, the choice to quit higher education voluntarily is not controlled by the institution itself. Instead, the factors that lead to this decision are more voluntary in nature, although students may still to a certain degree be unwilling to leave university (for example, in the case of health reasons or indigence).

2.3.3.5 Caveats

Despite exploiting many factors that impact the dependent variable in the models, this work does not establish the causal effect of ethnicity on the likelihood of graduation/withdrawal. The reason is that ethnicity is likely endogenous, in the sense that several unobserved characteristics may be correlated with both ethnicity and the response variable. In this case, the coefficient estimates for each ethnic group would be biased. Examples of omitted variables include students' learning styles, cultural attitudes towards higher education, the "wrong" choice of the subject of study, individual aspiration or self-motivation, the sense of "belongingness" to the institution, discrimination in assessments, and so forth. For instance, if the learning styles vary systematically between ethnic groups (Ridley, 2007), omitting this unobserved variable might bias the coefficients on ethnicity.

There might also be differential selection into the four types of institutions by ethnicity, which is not fully captured by the variables employed in the present analysis (that is, the prior educational attainment, the distance travelled, and the socio-economic background). For example, there are fewer Russell Group universities in the Midlands, where the proportion of some ethnic minorities is higher than in other UK regions (except London, where the competition to enter a prestigious university is more intense). Therefore, given that ethnic minorities are less likely to relocate than White students (Christie, 2007; Khambhaita and Bhopal, 2015), they might have fewer opportunities to attend an elite university.

Although I account for several university quality measures, the current datasets do not contain other institution-related characteristics (such as university structures, the level of academic support, and pastoral care). One way to address this issue would be to include university fixed effects in the MLR models. However, apart from the fact that this approach is too costly from a computational perspective, the existing software commands (Pforr, 2014) do not allow using marginal effects after executing the fixed-effects MLR models. Also, these commands are not compatible with interaction terms, thus rendering this method disadvantageous in practice.

Moreover, some explanatory variables are likely on the "causal path" of ethnicity. For example, if there is a causal effect of ethnicity on prior educational attainment (because of institutional factors in the pre-university years), then examining the impact of ethnic minority groups on the likelihood of graduation/dropout would be problematic when the models contain the *tariff score* variable (which would thus be a "bad control"). Specifically, controlling for *tariff score* implies that the treatment group comprises ethnic minority students attaining the same level of prior achievement as White students. Hence, these non-White students may be above average "ability". Therefore, adjusting the models for prior attainment (and assuming that ethnicity lowers attainment) would compare more able non-White students with less able White undergraduates, thus underestimating the true ethnic gap in the completion probability. This should be particularly relevant for specific ethnic minorities, such as Black, Pakistani, and Bangladeshi students, who perform significantly worse than the other ethnic groups before entering higher education (as shown in Figure 2.2).

Finally, the information about the dependent variable ("reason for ending course") in the HESA datasets is recorded by universities when they close a student instance. However, there might be instances in the data that two outcomes occur at the same time. For example, some students may be close to failing exams and decide to leave before university forces them to quit. This could be problematic methodologically if these instances cover a substantial proportion of the sample or if this proportion varies significantly across ethnic groups. Given that I cannot investigate the above hypothesis, I assume that the dependent variable captures the actual reason for students leaving university in the present paper.

2.4 Results

This section deciphers the relationship between the probability of graduation/non-completion and all the explanatory variables comprised in the multinomial logistic regression, based on the average adjusted predictions and the respective average marginal effects described earlier. I primarily focus on ethnicity, which is the key independent variable of interest. Moreover, I present and interpret the interactive effects between ethnicity and gender, type of institution, socio-economic background, and prior attainment.

2.4.1 Average marginal effects and adjusted predictions

The findings indicate that controlling for a plethora of factors in the regression models reduces, but does not eliminate, the ethnic gaps in the probability of academic dismissal. The results show that all ethnic minority groups are more likely to fail their degrees than White students (Table 2.5). Specifically, the ethnic differences in the average probability of academic failure range from 1.2 percentage points for the Indian and Chinese students to 3.3 percentage points for the Black African undergraduates, keeping all else constant. The likelihood of failing academically is exceptionally high for Black African students (7.1%), nearly twice as large as that of White undergraduates (Table 2.6 and Figure 2.3).

Whilst White students perform better in terms of academic assessments than their peers from ethnic minority backgrounds, this picture is reversed when examining the voluntary dropout propensities. Specifically, the probability of leaving university voluntarily for any other reasons is higher for White students, with the differences ranging from half percentage point in favour of students from Mixed ethnic backgrounds to 2.1 percentage points in favour of the Black African group.

However, in absolute terms, the ethnic disparities in voluntary non-completion rates are smaller than those in compulsory withdrawal (see Figure 2.4), suggesting that academic failure is the primary cause of discrepancies in student attrition. As a corollary, White students experience, on average, a higher propensity to complete their undergraduate studies than most ethnic minorities, except for Indian and Chinese groups (Table 2.6). In particular, the graduation likelihood is identical for White, Chinese, and Indian first-degree students (89.4%-

89.6%), but it is statistically significantly larger than that of the rest ethnic groups, notably with respect to the Black community (88.3%-88.4%).

Analysing the effects of the other explanatory variables on non-completion reveals some interesting findings (Table 2.5). Men are 1.3 percentage points less likely to complete their degrees than women, which is principally influenced by their higher chances of failing academically. This confirms the previous evidence that males' performance across all education phases (from primary school to university) is worse than that of females (Hillman and Robinson, 2016). Therefore, the higher probability of men failing to navigate higher education and meet the university's academic standards may reflect gender discrepancies traced back in earlies stages of the education system, partially explained by different learning style preferences between genders (OECD, 2015).

It is noteworthy that the HE participation rate is higher for females than for males across all social classes and ethnic groups (Crawford and Greaves, 2015). In particular, the percentage of females accessing university started rising at the end of the 1970s. It surpassed the respective men's share in 1992 for the first time and it has progressively improved since then (Bowes et al., 2015). The men's disadvantage regarding school attainment and the chances of accessing higher education are more pronounced among White people from low socio-economic backgrounds than others (House of Commons, 2014; BIS, 2016).

The present results show that prior educational attainment (measured by the tariff score) is negatively related to both voluntary dropout and involuntary withdrawal, but its effect is more potent in the latter case. The students who attended private schools before university are more likely to graduate than those from state schools, keeping all else equal. As expected, family background (measured by parental occupation and education) is a critical determinant of students' success in university. Specifically, undergraduates whose parents work in higher managerial/professional jobs or hold a university qualification have a greater probability of passing all their assessments and attaining their degree than others. However, students' socio-economic background is not significant in predicting voluntary withdrawal.

Non-completion rates differ markedly across subjects of study. Students in medicine & dentistry, and veterinary sciences are, by far, more likely to fail their assessments than those reading for other degrees. This might be attached to the fact that those courses take more years to complete, and programme duration is positively associated with the academic failure likelihood. Undergraduates studying humanities, arts, education, and mass communications are more likely to voluntarily leave higher education, although their probability of failure is lower than others. The pattern in voluntary dropout for the latter courses is probably connected with opportunity costs, as it is well-established from the UK literature that STEM and LEM subjects offer a higher wage premium in the labour market relative to other degrees (Britton et al., 2016; Walker and Zhu, 2018; Belfield et al., 2018a, 2018b).

Parenthetically, survey-based research has found that the wrong choice of study is a salient withdrawal cause cited by students, leading to reduced motivation, which, in turn, affects their academic performance and the dropout likelihood (Christie, Munro and Fisher, 2004). Therefore, comprehensive guidance and support from parents and schools (in collaboration with the higher education providers) can enhance the students' knowledge regarding the courses and universities, thus helping them align their pre-entry expectations with their actual higher education experience. In this context, policy initiatives have focused on improving the information provision to prospective students before applying to an institution, for example, through outreach programmes, "open days", and other activities (BIS, 2014).

The probability of involuntary and voluntary dropout is more remarkable for part-time enrollees, which implies that these students have increased difficulties in combining studies with work or other commitments. Although part-time students cover a small fraction of young undergraduates' sample, their proportion among ethnic minorities is more than twice that among White students (5.3% versus 2.3%).

Students attending the well-regarded Russell Group universities choose to quit willingly more often than those in other pre-1992 and post-1992 institutions, but they are less likely to fail academically. Taken together, the average probability of graduation differs little among these three institution types, keeping all else fixed. This mixed picture might explain why previous studies did not direct their attention to exploring differences in the dropout rates among university types, as they could not distinguish between the different reasons for withdrawal (e.g., Crawford, 2014b). Undergraduates studying in London perform worse than those in all other England and Wales regions in terms of completion likelihood. The "grey" literature has identified this London's drawback in university retention rates, relating it to issues around poor preparation and support, students' channelling into "wrong" courses because of ineffective advice before entry, living costs, and mental health problems (SMF, 2017; Petrie and Keohane, 2019).

The results presented here confirm a strong association between peer effects and the likelihood of non-completion (see also Johnes and McNabb, 2004; Collings, Swanson and Watkins, 2014). On average, a higher proportion of non-White peers on a course reduces the average probability of quitting voluntarily, while it enhances the chances of successful completion, other things equal. There is also a differential effect of students' ability mismatch (that is, a person's ability relative to his/her fellows' average capabilities) on student retention according to the non-completion types. Specifically, the students' previous attainment as a fraction of their peers' average (that is, the relative tariff score) is positively correlated with voluntary dropout, but it is negatively correlated with failure. Consistent with the findings of Johnes and McNabb (2004), the university selectivity (measured by the average institution's tariff score) escalates the non-completion probability, primarily through the likelihood of academic dismissal.

Examining the effect of other university attributes on non-completion leads to divergent conclusions. While the level of university income per student is positively related to the graduation probability, it also aggravates the voluntary dropout propensity, keeping all else equal. There are also diverse effects of the institutions' TEF levels on student retention, depending on the response variable's outcome. Furthermore, a higher academic staff-student ratio boosts completion rates and is inversely correlated with voluntary leave, although it positively impacts the likelihood of failure. As Johnes and McNabb (2004) note, this mixed picture might indicate that an increased staff-student ratio possibly echoes the pastoral feature of institutional support, instead of academic support.

Finally, the model outcomes show that increasing the non-White/White academic staff ratio would substantially ameliorate both voluntary and involuntary dropout rates. Improving ethnic diversity through the latter ratio should become particularly beneficial for ethnic minority students, as it could raise their sense of belongingness to the institution, while it might also strengthen their motivation to follow an academic career (UUK and NUS, 2019).

Table 2.5. Multinomial logistic regression: Average marginal effects (AMEs)
Dependent variable: Reason for ending course

Variable	Successful completion		Academic failure		Other reason	
	AMEs	SE	AMEs	SE	AMEs	SE
Ethnic group						
White	+	+	+	+	+	+
Black Caribbean	-0.010***	(0.002)	0.025***	(0.002)	-0.015***	(0.002)
Black African	-0.011***	(0.001)	0.033***	(0.001)	-0.021***	(0.001)
Other Black	-0.011**	(0.004)	0.025***	(0.004)	-0.014***	(0.004)
Indian	0.002**	(0.001)	0.012***	(0.001)	-0.014***	(0.001)
Pakistani	-0.008***	(0.001)	0.018***	(0.001)	-0.010***	(0.001)
Bangladeshi	-0.008***	(0.003)	0.021***	(0.003)	-0.014***	(0.002)
Chinese	0.001	(0.002)	0.012***	(0.002)	-0.013***	(0.002)
Other Asian	-0.006***	(0.002)	0.021***	(0.002)	-0.014***	(0.002)
Mixed	-0.009***	(0.001)	0.013***	(0.001)	-0.005***	(0.001)
Other ethnic group	-0.008***	(0.002)	0.023***	(0.002)	-0.015***	(0.002)
Student's characteristics						
Male	-0.013***	(0.000)	0.016***	(0.000)	-0.003***	(0.000)
Age	0.006***	(0.000)	-0.000	(0.000)	-0.006***	(0.000)
Disability	-0.009***	(0.001)	0.001**	(0.001)	0.008***	(0.001)
Home fees eligible	-0.013***	(0.002)	-0.001	(0.002)	0.014***	(0.002)
Socio-economic background (parental occupation)						
Higher managerial/professional	+	+	+	+	+	+
Lower managerial/professional	-0.002***	(0.001)	0.001**	(0.001)	0.000	(0.001)
Intermediate	-0.002***	(0.001)	0.002**	(0.001)	0.000	(0.001)
Small employers/own account workers	-0.002**	(0.001)	0.002**	(0.001)	0.000	(0.001)
Technical/lower supervisory	-0.004***	(0.001)	0.003***	(0.001)	0.001	(0.001)
Semi-routine	-0.005***	(0.001)	0.004***	(0.001)	0.000	(0.001)
Routine	-0.006***	(0.001)	0.005***	(0.001)	0.000	(0.001)
Long-term unemployed/Never worked	-0.006	(0.004)	0.004	(0.003)	0.002	(0.003)
Unknown	-0.004***	(0.001)	0.003***	(0.001)	0.000	(0.001)
Parental education						
Parents without HE qualifications	+	+	+	+	+	+
Parents with HE qualifications	0.003***	(0.000)	-0.001**	(0.000)	-0.002***	(0.000)
Unknown/Refused	0.000	(0.000)	0.003***	(0.000)	-0.003***	(0.000)
University type						
Russell Group	+	+	+	+	+	+
Other pre-1992	0.001	(0.001)	0.018***	(0.001)	-0.020***	(0.001)
Post-1992	0.002*	(0.001)	0.021***	(0.001)	-0.023***	(0.001)
Specialist	-0.020***	(0.002)	0.022***	(0.002)	-0.002	(0.002)
Region of university						
London	+	+	+	+	+	+
North East	0.044***	(0.001)	-0.030***	(0.001)	-0.014***	(0.001)
North West	0.005***	(0.001)	-0.016***	(0.001)	0.011***	(0.001)
Yorkshire	0.011***	(0.001)	-0.019***	(0.001)	0.009***	(0.001)
East Midlands	0.024***	(0.001)	-0.029***	(0.001)	0.005***	(0.001)
West Midlands	0.011***	(0.001)	-0.023***	(0.001)	0.012***	(0.001)
East of England	0.018***	(0.001)	-0.025***	(0.001)	0.007***	(0.001)
South East	0.008***	(0.001)	-0.028***	(0.001)	0.020***	(0.001)
South West	0.021***	(0.001)	-0.032***	(0.001)	0.011***	(0.001)
Wales	0.013***	(0.001)	-0.023***	(0.001)	0.009***	(0.001)
Scotland	-0.011***	(0.002)	-0.020***	(0.002)	0.031***	(0.002)
N. Ireland	-0.053***	(0.004)	0.083***	(0.005)	-0.030***	(0.002)

Continued on next page

Table 2.5. (continued)						
Subject of Study						
Social studies	+	+	+	+	+	+
Medicine & dentistry	-0.033***	(0.003)	0.054***	(0.004)	-0.021***	(0.003)
Subjects allied to medicine	-0.001	(0.001)	0.001	(0.001)	-0.000	(0.001)
Biological sciences	0.001	(0.001)	0.001	(0.001)	-0.002**	(0.001)
Veterinary science	-0.025**	(0.012)	0.049***	(0.016)	-0.024***	(0.008)
Agriculture & related subjects	-0.005**	(0.002)	0.005**	(0.003)	-0.001	(0.002)
Physical sciences	-0.009***	(0.001)	0.009***	(0.001)	0.000	(0.001)
Mathematical sciences	-0.018***	(0.001)	0.015***	(0.002)	0.002*	(0.002)
Computer science	-0.009***	(0.001)	0.009***	(0.001)	0.001	(0.001)
Engineering & technology	-0.010***	(0.001)	0.013***	(0.001)	-0.003***	(0.001)
Architecture, building & planning	-0.008***	(0.001)	0.005***	(0.001)	0.003**	(0.001)
Law	-0.001	(0.001)	0.001	(0.001)	-0.001	(0.001)
Business & administrative studies	-0.000	(0.001)	0.001	(0.001)	-0.000	(0.001)
Mass communications & documentation	-0.001	(0.001)	-0.004***	(0.001)	0.005***	(0.001)
Languages	-0.003***	(0.001)	-0.002*	(0.001)	0.005***	(0.001)
Historical & philosophical studies	0.000	(0.001)	-0.003***	(0.001)	0.003***	(0.001)
Creative arts & design	-0.000	(0.001)	-0.006***	(0.001)	0.006***	(0.001)
Education	0.001	(0.001)	-0.009***	(0.001)	0.008***	(0.001)
Combined	-0.003***	(0.001)	0.001	(0.001)	0.002***	(0.001)
Mode of study						
Part-time	+	+	+	+	+	+
Full-time	0.109***	(0.003)	-0.086***	(0.002)	-0.023***	(0.002)
Sandwich	0.105***	(0.003)	-0.088***	(0.002)	-0.017***	(0.002)
Other	0.093***	(0.003)	-0.083***	(0.003)	-0.011***	(0.002)
Length of programme						
<= 2 years	+	+	+	+	+	+
2-3 years	-0.064***	(0.002)	0.025***	(0.001)	0.039***	(0.002)
3-4 years	-0.073***	(0.002)	0.031***	(0.001)	0.042***	(0.002)
4-20 years	-0.080***	(0.002)	0.031***	(0.002)	0.049***	(0.002)
Year of student on course						
1st year	+	+	+	+	+	+
2nd year	0.308***	(0.003)	-0.033***	(0.003)	-0.274***	(0.004)
3rd year	0.842***	(0.003)	-0.270***	(0.004)	-0.572***	(0.004)
4rth and over	0.848***	(0.003)	-0.276***	(0.004)	-0.572***	(0.004)
Major source of tuition fees						
No award/backing	+	+	+	+	+	+
UK LEA award	0.014***	(0.001)	0.005***	(0.000)	-0.019***	(0.001)
Provider waiver	-0.007**	(0.003)	0.022***	(0.003)	-0.016***	(0.003)
UK central government	0.012***	(0.001)	-0.006***	(0.001)	-0.006***	(0.001)
Other	-0.009***	(0.002)	0.001	(0.001)	0.008***	(0.002)
Term-time accommodation						
Own residence	+	+	+	+	+	+
Parental/guardian home	0.001	(0.001)	-0.000	(0.001)	-0.001	(0.001)
Provider's property	-0.005***	(0.001)	0.003***	(0.001)	0.002**	(0.001)
Private-sector halls	-0.011***	(0.001)	0.009***	(0.001)	0.002**	(0.001)
Other rented	0.000	(0.001)	0.007***	(0.001)	-0.007***	(0.001)
Other	0.004***	(0.001)	0.004***	(0.001)	-0.008***	(0.001)
Not in attendance	0.000	(0.004)	-0.000	(0.005)	-0.000	(0.005)
Unknown	-0.000	(0.001)	0.002**	(0.001)	-0.002*	(0.001)
Pre-entry characteristics						
Tariff Score	0.0001***	(0.000)	-0.0001***	(0.000)	-0.000**	(0.000)
Public school	+	+	+	+	+	+
Private school	0.002***	(0.001)	-0.000	(0.001)	-0.002***	(0.001)
Unknown school type	0.004***	(0.001)	-0.004***	(0.001)	0.000	(0.001)
Distance travelled (km)	0.000*	(0.000)	-0.00002***	(0.000)	0.00002***	(0.000)

Continued on next page

Table 2.5. (continued)						
Peer effects and other university quality measures						
Relative tariff score	0.001	(0.002)	-0.004***	(0.002)	0.003**	(0.002)
Proportion of non-White peers	0.007***	(0.001)	0.015***	(0.001)	-0.022***	(0.001)
University's average tariff score	-0.0001***	(0.000)	0.0001***	(0.000)	-0.000***	(0.000)
Staff-student ratio	0.015***	(0.005)	0.016***	(0.006)	-0.031***	(0.006)
Non-White/White staff ratio	0.139***	(0.004)	-0.105***	(0.003)	-0.034***	(0.004)
University's income per student	0.001***	(0.000)	-0.001***	(0.000)	0.000***	(0.000)
TEF-Gold	+	+	+	+	+	+
TEF-Silver	0.001*	(0.000)	0.004***	(0.000)	-0.005***	(0.000)
TEF-Bronze	-0.008***	(0.001)	-0.004***	(0.001)	0.013***	(0.001)
TEF-Unknown	0.038***	(0.001)	-0.019***	(0.001)	-0.019***	(0.001)
Academic year						
2010/11	+	+	+	+	+	+
2011/12	-0.001**	(0.001)	-0.001	(0.001)	0.002***	(0.001)
2012/13	-0.002***	(0.001)	-0.003***	(0.001)	0.005***	(0.001)
2013/14	-0.003***	(0.001)	0.001	(0.001)	0.002***	(0.001)
2014/15	-0.005***	(0.001)	0.001*	(0.001)	0.004***	(0.001)
Observations	1,143,576					
Pseudo R^2	0.563					

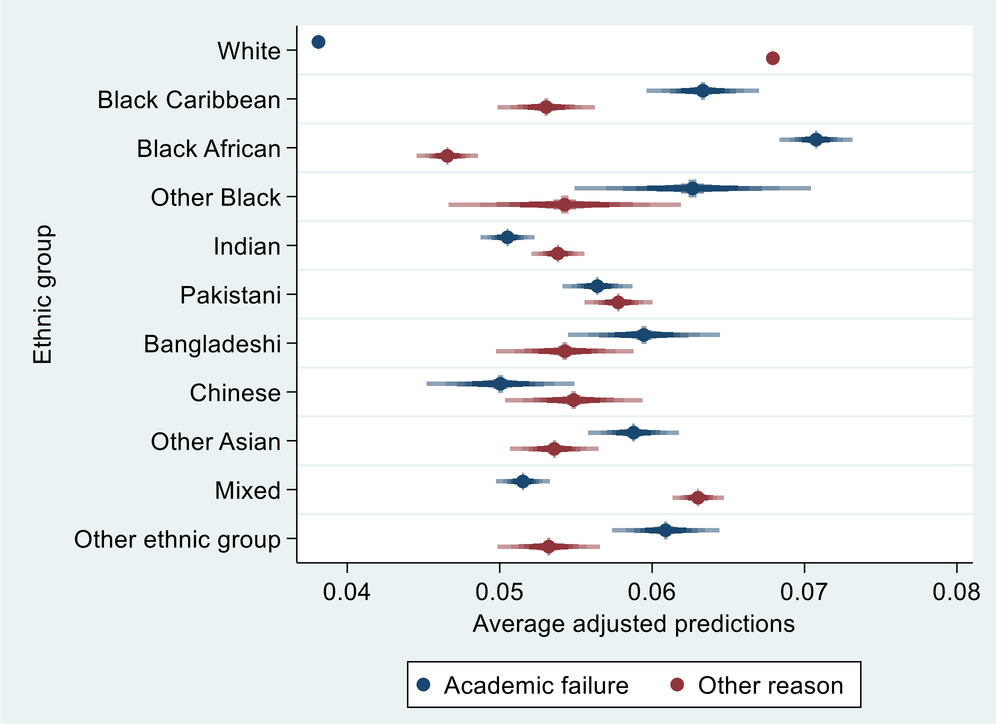
Note: The marginal effects shown in the table are derived from the original multinomial logistic regression (post-estimates). AMEs sum up to zero across all three outcomes.
SE: The standard errors of the AMEs are estimated based on the delta method. In the multinomial logistic regression, robust standard errors are used (that is, the variance estimator is robust to certain misspecification forms).
+Reference category.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
The "Other reason" outcome refers to voluntary dropout and includes: "health reasons", "death", "financial reasons", "other personal reasons", "written off after lapse of time", "exclusion", "gone into employment", and "other".
*The model includes the following independent variables and interaction terms: "ethnicity", "male", "age", "disability", "home fees eligible", "socio-economic classification", "parental education", "type of university", "region of university", "subject of study", "mode of study", "length of programme", "year of student on course", "major source of tuition fees", "term-time accommodation", "tariff score", "type of school", "distance travelled", "relative tariff score", "proportion of non-White peers", "university's average tariff score", "staff-student ratio", "non-White/White staff ratio", "university's income per student", "TEF award", "academic year", "ethnicity*male", "ethnicity*type of university", "ethnicity*socio-economic classification", "ethnicity*tariff score", and "relative tariff score squared".*
The cases with unknown ethnicity (<1% of the initial sample) are dropped from the regression analysis. For the dummy variables with a significant proportion of missing (unknown) values (>5%), I have included an additional category ("unknown").
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.6. Average adjusted predictions (AAPs)

	Dependent variable: Reason for ending course		
Ethnic group	**Successful completion**	**Academic failure**	**Other reason**
White	0.894***	0.038***	0.068***
	(0.000)	(0.000)	(0.000)
Black Caribbean	0.884***	0.063***	0.053***
	(0.002)	(0.002)	(0.002)
Black African	0.883***	0.071***	0.047***
	(0.001)	(0.001)	(0.001)
Other Black	0.883***	0.063***	0.054***
	(0.004)	(0.004)	(0.004)
Indian	0.896***	0.051***	0.054***
	(0.001)	(0.001)	(0.001)
Pakistani	0.886***	0.056***	0.058***
	(0.001)	(0.001)	(0.001)
Bangladeshi	0.886***	0.059***	0.054***
	(0.002)	(0.003)	(0.002)
Chinese	0.895***	0.050***	0.055***
	(0.002)	(0.002)	(0.002)
Other Asian	0.888***	0.059***	0.054***
	(0.001)	(0.002)	(0.001)
Mixed	0.885***	0.052***	0.063***
	(0.001)	(0.001)	(0.001)
Other ethnic group	0.886***	0.061***	0.053***
	(0.002)	(0.002)	(0.002)
Observations		1,143,576	
Pseudo R^2		0.563	

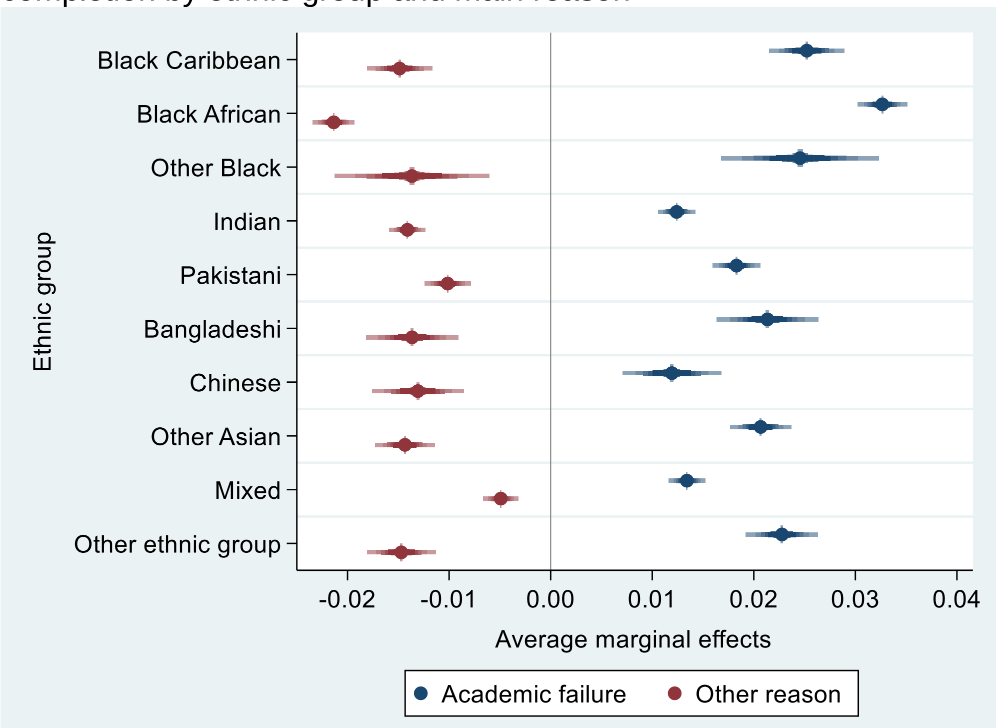
Note: The predicted probabilities shown in this table are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. AAPs sum up to one (100%) across all three outcomes.
Standard errors in parentheses based on the delta method. In the multinomial logistic regression, robust standard errors are used (that is, the variance estimator is robust to certain misspecification forms).
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*
The "Other reason" outcome refers to voluntary dropout and includes: "health reasons", "death", "financial reasons", "other personal reasons", "written off after lapse of time", "exclusion", "gone into employment", and "other".
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.3. Average adjusted predictions by ethnic group and main dropout reason (probability of non-completion)



Note: For each ethnic group, the figures represent the average probability of academic failure and voluntary dropout ("other reason") conditional on the total pool of observed characteristics (see Table 2.6). The shaded areas represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.4. Average marginal effects (AMEs) on the probability of non-completion by ethnic group and main reason



Note: For each ethnic minority group, the figures (AMEs) represent the difference in the conditional likelihood of academic failure and voluntary dropout ("other reason") relative to White undergraduates (reference category). The shaded areas represent the 95% confidence intervals of the estimated marginal effects.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

2.4.2 Interaction effects

This subsection presents the interaction effects between students' ethnicity and gender, social class, type of university attended, and prior attainment on the probability of failing their degree, based on the "second differences" approach. As noted earlier, I concentrate on the interaction effects concerning academic failure, as the ethnic disparities are more prominent within this dependent variable's outcome.

I begin with the gender aspect of interdependence (Table 2.7). The results show statistically significant gender differences in the likelihood of academic failure among students from Black, Pakistani, Bangladeshi, and Mixed ethnic backgrounds. Specifically, the differences in the probability of involuntary attrition within each of these groups (relative to the White majority group) are smaller for women than men, with the respective gender gap ranging between 0.6 and 0.9 percentage points. In contrast, there are no gender differences in the likelihood of involuntary non-completion between the rest of ethnic groups and their White counterparts. In other words, although, for example, Indian and Chinese undergraduates are more likely to drop out due to academic failure than their White counterparts of similar observed characteristics (see also Figure 2.5), the size of these gaps does not depend on their gender.

There is a substantial interrelationship between ethnicity and the university type in the probability of academic failure (Table 2.8 and Figure 2.6). On average, all ethnic minority students perform better in the Russell Group institutions than in the rest of higher education providers in terms of their likelihood of dropping out because of academic failure. For instance, the ethnic gap in the failure propensity for an average Black African undergraduate enrolled at a Russell Group university (relative to his/her White counterparts) is 2.6 percentage points lower than that of a Black African fellow attending a post-1992 university. Previous research indicates that ethnic minorities are under-represented in "old" universities (Boliver, 2013, 2016). The rising political and social pressure for offering equal opportunities has likely encouraged the Russell Group and other "old" institutions to enhance their structures, activities and student support services (Russell Group, 2019) aiming to improve the ethnic minorities' university experience, which has, in turn, resulted in a decrease in students' performance gaps.

Table 2.9 shows how the effect of ethnicity on academic failure likelihood changes across selected levels of undergraduates' socio-economic background (see also Figure 2.7). I recognise that the small sample size for some subgroups of students may have compromised the statistical significance of the respective interaction effects (especially for the "Other Black" category). Nonetheless, the results presented here do not provide evidence that the magnitude of ethnic gaps in academic failure differs substantially across various segments of the socio-economic distribution (for most ethnic minorities). In two of the few exceptions, the ethnic difference in the probability of involuntary dropout exacerbates for Black African and "Other Asian" students (compared to their White peers) as we move from the highest to lower social class distribution levels. More specifically, this ethnic gap for Black African students almost doubles from 2.5 percentage points for those with parents in higher managerial/professional occupations (such as chief executives, lawyers, large employers, university lecturers, and general practitioners) to 4.5 percentage points for those in routine professions (such as machine operators, motor vehicle drivers, and cleaners).

Finally, Figure 2.8 illustrates how the likelihood of academic failure varies across chosen values[9] of prior academic ability (measured by the *tariff score* variable). As expected, for all ethnic groups, the probability of failing at university declines as the level of prior attainment increases, keeping all else equal. Nevertheless, the slope of the curve is not constant for all ethnic groups, confirming that there are heterogeneous effects of prior attainment on the propensity of involuntary attrition. For instance, the curve declines more sharply for Pakistani undergraduates (indicated by the bright red line) than others as we move to higher levels of academic ability. This suggests that the ethnic differences in the failure likelihood (relative to White students) are more salient for low-ability Pakistani students than their high-ability Pakistani peers. Therefore, to tackle the ethnic gaps in academic performance, universities should ensure that students with a lower educational profile on entry receive the necessary study skills (for example, through targeted training sessions and support), especially at the beginning of their university life.

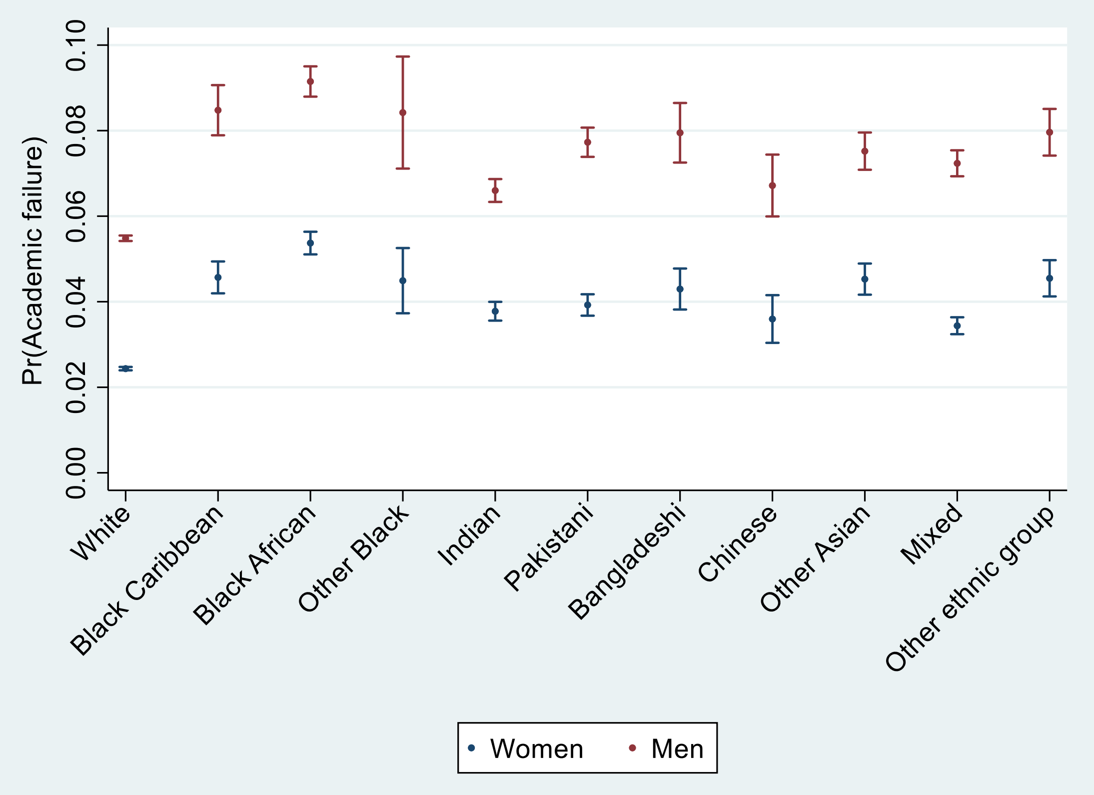
[9] *Because* tariff score *is a continuous variable, Figure 2.8 presents specific values of its distribution to explore the interactive effect of "ethnicity*tariff score" on academic dismissal likelihood. These values correspond to representative points of interest, which also have many observations. For instance, for 97% of graduates, the tariff score value is less than 600 points, whereas approximately one-third (34%) scored below 300 points. The mean value of the tariff score in the regression sample is 347.*

Table 2.7. Interaction effects of gender and ethnicity on the probability of academic failure

Ethnic group	Ethnic gap in Pr(Academic failure)		Second difference
	Women	Men	
Black Caribbean	0.021***	0.030***	-0.009***
Black African	0.029***	0.037***	-0.007***
Other Black	0.021***	0.029***	-0.009
Indian	0.013***	0.011***	0.002
Pakistani	0.015***	0.022***	-0.008***
Bangladeshi	0.019***	0.025***	-0.006*
Chinese	0.012***	0.012***	-0.001
Other Asian	0.021***	0.020***	0.001
Mixed	0.010***	0.018***	-0.007***
Other ethnic group	0.021***	0.025***	-0.004

Note: The coefficients within each gender (ethnic gap) represent the difference in the average probability of dropping out due to academic failure between each ethnic minority and White students (reference category). The rightmost column shows the difference in the ethnic gap between men and women (i.e., second difference). The results are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5.
** p < 0.1, ** p < 0.05, *** p < 0.01*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.5. Probability of academic failure by gender and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. The error bars represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.8. Interaction effects of university type and ethnicity on the probability of academic failure

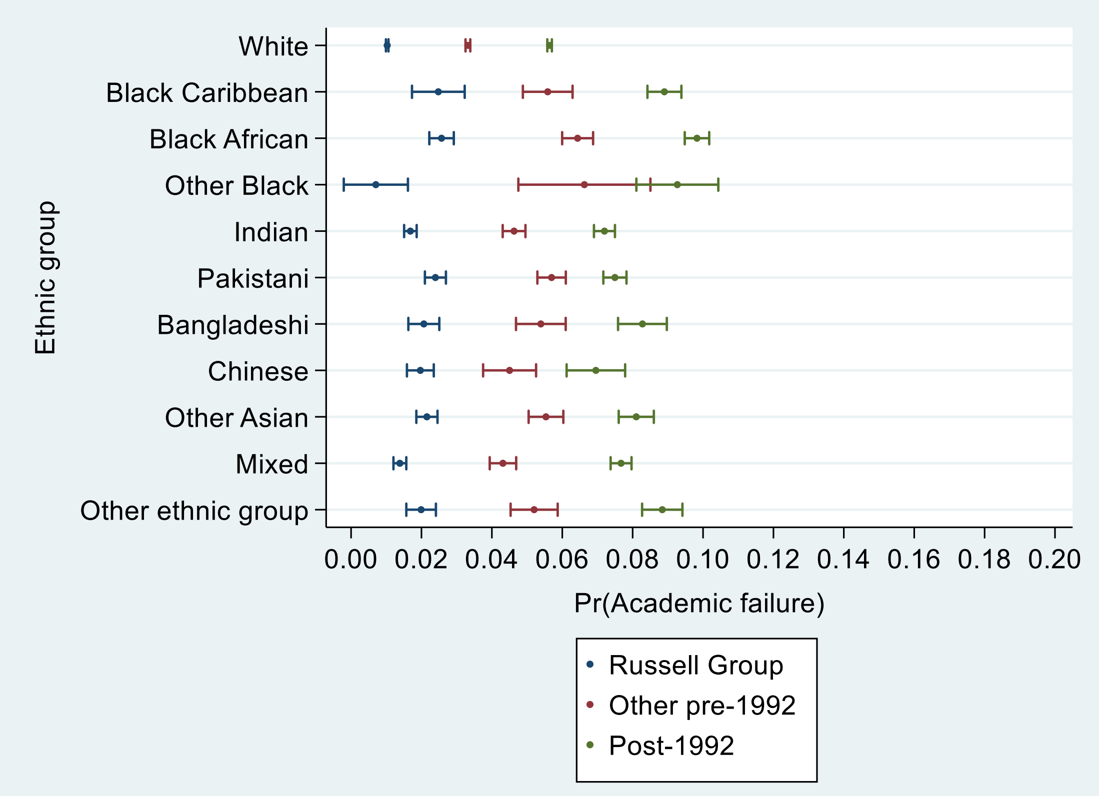
Ethnic group	Ethnic gap in Pr(Academic failure)			Second difference		
	Russell Group (1)	Other pre-1992 (2)	Post-1992 (3)	(1-2)	(1-3)	(2-3)
Black Caribbean	0.015***	0.023***	0.033***	-0.008	-0.018***	-0.010**
Black African	0.015***	0.031***	0.042***	-0.016***	-0.026***	-0.011***
Other Black	-0.003	0.033***	0.036***	-0.036***	-0.040***	-0.003
Indian	0.007***	0.013***	0.016***	-0.007***	-0.009***	-0.002
Pakistani	0.014***	0.024***	0.019***	-0.010***	-0.005**	0.005**
Bangladeshi	0.010***	0.021***	0.026***	-0.010***	-0.016***	-0.006
Chinese	0.009***	0.012***	0.013***	-0.002	-0.004	-0.001
Other Asian	0.011***	0.022***	0.025***	-0.011***	-0.013***	-0.002
Mixed	0.004***	0.010***	0.020***	-0.006***	-0.017***	-0.010
Other ethnic group	0.010***	0.019***	0.032***	-0.009**	-0.022***	-0.013

Note: The coefficients within each university type (ethnic gap) represent the difference in the average probability of dropping out due to academic failure between each ethnic minority and White students (reference category). The "second difference" columns show the difference in the ethnic gap between type of institutions. The results are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. Specialist institutions are not presented in the table as they cover a small proportion of students (1.4%).

** p < 0.1, ** p < 0.05, *** p < 0.01*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.6. Probability of academic failure by university type and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. The error bars represent the 95% confidence intervals of the estimated predicted probabilities. Specialist institutions are not presented in the graph, as they cover a small proportion of students (1.4%) and their wide confidence intervals overlap with the rest categories.

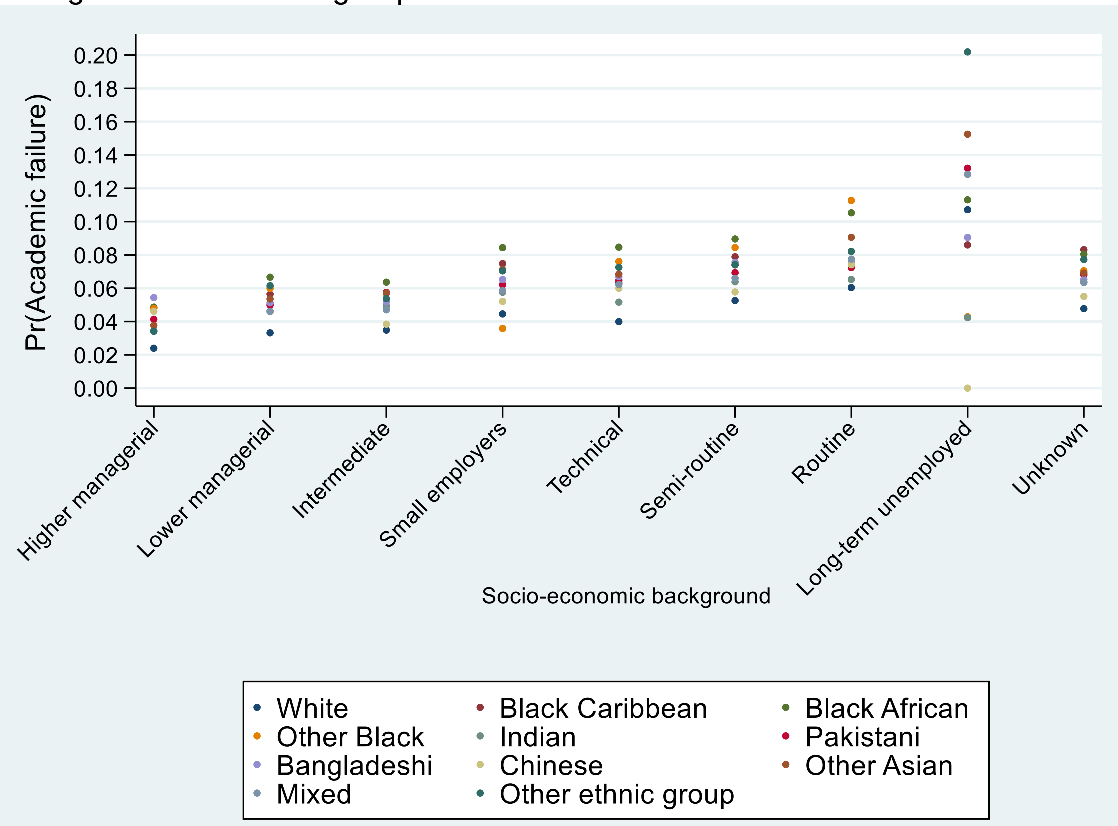
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.9. Interaction effects of socio-economic background and ethnicity on the probability of academic failure

Ethnic group	Ethnic gap in Pr(Academic failure)			Second difference		
	Higher managerial (1)	Small employers (2)	Routine (3)	(1-2)	(1-3)	(2-3)
Black Caribbean	0.024***	0.030***	0.014*	-0.006	0.010	0.016
Black African	0.025***	0.040***	0.045***	-0.015**	-0.020***	-0.005
Other Black	0.024***	-0.009	0.052***	0.033	-0.028	-0.061**
Indian	0.010***	0.013***	0.005	-0.003	0.006	0.008
Pakistani	0.017***	0.018***	0.012***	0.000	0.005	0.006
Bangladeshi	0.030***	0.021***	0.015**	0.010	0.015	0.005
Chinese	0.022***	0.008	0.014	0.015*	0.009	-0.006
Other Asian	0.014***	0.027***	0.030***	-0.013**	-0.016**	-0.004
Mixed	0.010***	0.014***	0.017***	-0.004	-0.007	-0.003
Other ethnic group	0.010***	0.026***	0.022***	-0.016**	-0.011	0.004

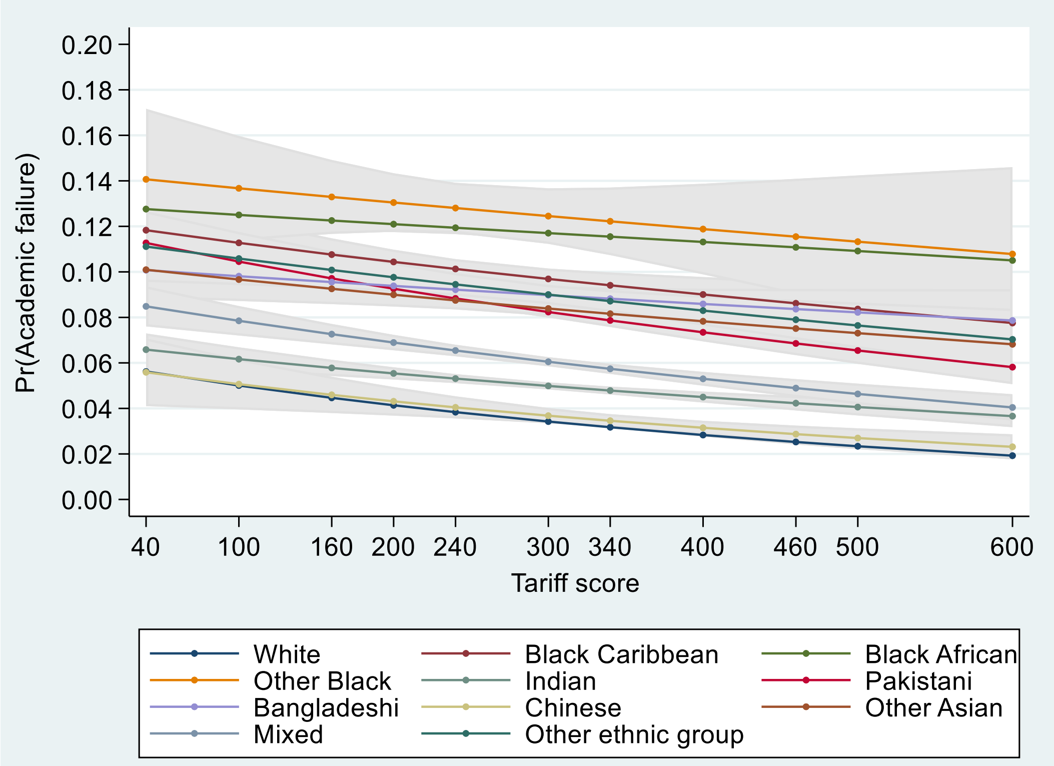
Note: The coefficients within each level of socio-economic background (ethnic gap) represent the difference in the average probability of dropping out due to academic failure between each ethnic minority and White students (reference category). The "second difference" columns show the difference in the ethnic gap between the socio-economic levels. The results are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5.
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.7. Probability of academic failure by level of socio-economic background and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. For clarity, the confidence intervals of the predicted probabilities are not plotted in the graph, as they overlap across ethnic groups.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 2.8. Probability of academic failure by ethnic group at selected levels of tariff score



Note: The predicted probabilities shown in the graph are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. The shaded areas represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

2.4.3 Robustness checks

I conducted certain robustness tests to evaluate the sensitivity of the main findings. First, I dropped the students who left university because of health reasons or death from the multinomial logistic regression sample. The reason is that these students are not considered "voluntary" dropouts in the precise sense of the term. Second, I incorporated the additional interaction term "*ethnicity*subject of study*" in the regression analysis to capture the potential differential effect of ethnicity on the probability of graduation, voluntary dropout, and compulsory withdrawal across various courses. Both robustness checks did not alter the results. As Table 2.A8 in the Appendix shows, the average marginal effects of each ethnic minority group remain impressively identical to those presented in the initial results (Table 2.5), even after encompassing the additional interaction term and excluding the students who did not complete their studies due to health reasons/death.

The third robustness check refers to how the interaction effects are technically calculated in practice when adopting the AMEs approach in multinomial logit models[10]. The results reported in subsection 2.4.2 (Tables 2.7-2.9) represent the interaction effect of specific explanatory variables (gender, university type, and social class) at different levels of ethnicity (that is, across ethnic groups). However, some authors recommend that one should also look at the other side of intersectionality by examining the interaction effect of ethnicity at various levels of the other independent variables (Mize, 2019). The results of this exercise, which are presented in Table 2.A9 of the Appendix, show that there are some differences in the magnitude of the interaction effects between the two approaches explained above (especially for students from “Other Black” backgrounds, probably because of their relatively small number of observations). However, these differences are minor across most ethnic groups, while the sign and statistical significance of the interaction effects remain unchanged in most cases.

2.5 Conclusion

Using recent individual-level data from the Higher Education Statistics Agency, this paper examines whether the likelihood of degree non-completion differs between undergraduates from diverse ethnic backgrounds in the UK higher education. For the first time in the ethnicity context, this study distinguishes between compulsory withdrawal (because of academic failure) and voluntary dropout (because of employment, personal, financial, or other reasons), recognising that the policy response to student withdrawal should be different depending on the dropout causes.

The multinomial logit model results provide firm evidence that, on average, all ethnic minority groups have a higher probability of failing their degrees than White students. Most worryingly, Black African students are twice as likely (7.1%) as their White peers (3.8%) to fail. On the contrary, White students have a higher propensity than ethnic minorities to quit voluntarily, although the differences are

[10] *For example, the first part of the Stata command I used to compute the interaction effect of gender across ethnic groups is “*margins, at (ethnicity) over (gender)*”. Before computing the predicted probabilities (and after producing the original regression coefficients), Stata splits the sample into two groups (males and females). Then, within each of these groups (as defined by the “*over()*” option of the command), Stata estimates the marginal effects at the specified values of ethnicity (as denoted by the* at () *option), which are treated as fixed. Subsequently, the second difference approach described in subsection 2.3.3.3 compares whether any two marginal effects differ in a statistically significant way. In the robustness check presented here, the Stata command changes to the other way round, that is, “*margins, at (gender) over (ethnicity)*”.*

smaller than those related to academic dismissal. The approach used in this work arguably explains why some earlier studies found contradictory results regarding the effect of ethnicity on the dropout likelihood, as they could not differentiate between voluntary and involuntary attrition (National Audit Office, 2007; Vignoles and Powdthavee, 2009). When considering heterogeneous effects, this paper shows that the ethnic gaps in academic failure (that is, the differences in the likelihood of failing the degree between each ethnic minority and the White students) are more pronounced for men than women and are less noticeable in the Russell Group universities relative to other institution types. In contrast, in most cases, the ethnic disparities in the probability of academic dismissal do not deviate significantly across various social class levels.

A central limitation of this work is that it is not feasible to control for all the determinants that influence the non-completion probability, which are likely to vary systematically across ethnic groups. These factors may relate to the poor choice of the subject of study, cultural attitudes and self-motivation, learning styles, university environment and available resources that affect the sense of “belongingness” to the institution, discrimination in teaching support and assessments, and other reasons concerning students’ social integration into university (Christie, Munro and Fisher, 2004; Yorke and Longden, 2008; Thomas et al., 2017). Therefore, the present study does not discern causality in the correlation between ethnicity and the likelihood of non-continuation.

However, this paper’s findings should improve the understanding by national policymakers and universities of determinants of dropout and help devise specific forms of support to students from different ethnic backgrounds. In particular, the extensive information about individual socio-demographic traits, prior attainment and university-related characteristics used in the econometric models does not fully explain the ethnic gaps in dropout rates. Therefore, while policymakers need to assess the relationship between these factors and degree completion likelihood as established in the present paper, they should also pay attention to identifying the possible unobserved determinants of student attrition mentioned above.

Interestingly, the results from the interaction effects analysis reveal a remarkable consistency in the rank ordering of ethnic gaps in the probability of academic failure across various university types and socio-economic groups, between genders and at different levels of prior ability. For instance, the ordering of the effect each ethnic minority has on the likelihood of involuntary dropout (relative to the White group) is similar within each type of higher education institution. Also, the ordering of university types is consistent across ethnic groups (see Figure 2.6). Thus, it seems rather unlikely that the ethnic gaps in the dropout likelihood are solely driven by specific unobserved characteristics, which are common across all ethnic minorities and different to White students. Instead, the consistent ordering of ethnic gaps plausibly reflects structural factors that transcend university type and explain why students of non-White ethnicity are more likely to drop out. From a policymaking viewpoint, it is therefore critical to identify the structural barriers that inhibit equity for ethnic minority students and affect their university experience. Such barriers may be associated with institutional culture and support systems, financial impediments related to tuition fees, rigid course content and delivery, inflexible formal assessments, and racial discrimination.

To date, policymakers have designed and implemented strategies that promote participation in higher education for individuals from disadvantaged backgrounds. For instance, one Government's commitment over the last years was to raise by 20% the number of ethnic minority people accessing higher education by 2020 relative to 2009 (BIS, 2016)[11]. However, less attention has been put towards retention of ethnic minority students and their success once they enrol at university. Therefore, the first policy recommendation would be to introduce a specific and measurable goal to tackle the ethnic gaps in the dropout propensity and track its progress within the next few years, particularly for Black undergraduates who are more likely than others to fail university.

At the university level, policy interventions should aim at developing ethnically inclusive environments by improving the offered support services (e.g., through mentoring programmes). These interventions could also focus on changing the "campus culture" by transforming the current climate and culture rather than merely addressing perceived issues with the ethnic minority students themselves. In addition, implementing learning analytics (while considering the respective ethical implications) would allow monitoring students' progress throughout their academic life stages and may improve student retention,

[11] *The number of "home" ethnic minority students with an accepted application via UCAS increased by 64% in 2020 (141 thousand accepted applicants) compared to 2010 (86 thousand) (Bolton, 2021).*

teaching and learning (Higher Education Commission, 2016). Adjusting the curriculum content and making the university attainment-gap data available to the academic staff might reduce ethnic discrepancies in attainment and identify students who need supplementary academic support. In the same context, it is fundamental to provide students whose entry profile is significantly lower than others with specific study skills, including, for example, compulsory training sessions tailored to the requirements of each course. This predominantly applies to Black, Pakistani, and Bangladeshi undergraduates, whose prior educational attainment is significantly lower than that of other ethnic groups.

Moreover, other countries have used performance-based funding schemes, which link institutions' funding to student retention, to incentivise universities to take actions that facilitate study success (Vossensteyn et al., 2015). Such funding instruments could be extended in the framework of eliminating ethnic inequalities associated with dropout rates. Given that ethnic minorities (especially the Black students) are more likely to encounter financial problems and, therefore, drop out from higher education, more universities should commit to financial support as another tool that would improve student retention. The financial support could take forms of connecting funding to progress (thus stimulating students to perform well and complete their studies) and additional university bursaries for undergraduates coming from low-income families. Moreover, developing more effective information campaigns at schools regarding the university and course options would likely enhance the matching between prospective students' academic abilities and the requirements of each university/subject of study. As a corollary, such policy actions might minimise the dropout rates stemming from the wrong choice of subject (SMF, 2017).

Further research is essential to comprehend better the impediments and complex mechanisms that undermine ethnic minorities' academic performance and higher education experience. Adopting a mixed-method approach by exploiting qualitative and quantitative data and methods will likely unveil the impact of hard-to-quantify variables (which are not observed in administrative datasets) on academic failure and the voluntary decision to leave university. This approach would require performing qualitative interviews with a representative group of university students and staff across various UK regions. The respective questionnaires should contain sections related to students' socio-demographic characteristics, learning styles, cultural attitudes and unfulfilled expectations, institutional structures, and racial discrimination.

Finally, exploring whether there are disproportionate ethnic effects of the COVID-19 pandemic on the university non-completion likelihood is another critical area for future research. The UK unemployment figures have substantially worsened during the pandemic crisis (ONS, 2021). From a theoretical angle, the net impact of an economic decline on dropout rates depends on two contradictory forces (Becker, 1994; Ghignoni, 2017). First, it can decrease the opportunity costs of higher education, thus reducing the number of students who leave the university to go into employment. Second, the deterioration of employment prospects may compromise students' motivation to complete their studies. Early reports show that ethnic minorities experience more detrimental effects of the pandemic on health, earnings, and employment than White people (Lally, 2020; Bracke et al., 2021). These inequalities could be followed through in ethnic minorities' academic performance, given that, as the present study confirms, socio-economic and health factors are strongly correlated with the probability of dropping out of higher education.

References of Chapter 2

Advance HE, 2020. *Equality+higher education - Students statistical report 2020.* York.

Agresti, A., 2013. *Categorical Data Analysis.* 3rd ed. New York: Wiley.

Ai, C. and Norton, E.C., 2003. Interaction terms in logit and probit models. *Economics Letters,* 80(1), pp. 123-129.

Amemiya, T., 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature,* 19(4), p. 1483.

Arrow, K.J., 1973. Higher education as a filter. *Journal of Public Economics,* 2(3), pp. 193-216.

Arulampalam, W., Naylor, R.A. and Smith, J.P., 2005. Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time-series analysis for UK university students. *Economics of Education Review,* 24(3), pp. 251-262.

Becker, G.S., 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy,* 70(5, Part 2), pp. 9-49.

Becker, G.S., 1994. *Human capital: A theoretical and empirical analysis, with special reference to education.* 3rd ed.: University of Chicago Press.

Bedard, K., 2001. Human Capital versus Signaling Models: University Access and High School Dropouts. *Journal of Political Economy,* 109(4), pp. 749-775.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y., 2018a. *The relative labour market returns to different degrees.* London: IFS (Institute for Fiscal Studies), (DFE-RR787).

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y., 2018b. *The impact of undergraduate degrees on early-career earnings.* London: IFS (Institute for Fiscal Studies), (DFE-RR808).

Bennett, R., 2003. Determinants of Undergraduate Student Drop Out Rates in a University Business Studies Department. *Journal of Further and Higher Education,* 27(2), pp. 123-141.

BIS, 2014. *National strategy for access and student success in higher education.* London: BIS (Department for Business Innovation and Skills).

BIS, 2016. *Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice.* London: BIS (Department for Business Innovation and Skills).

Boliver, V., 2013. How fair is access to more prestigious UK universities? *British Journal of Sociology,* 64(2), pp. 344-364.

Boliver, V., 2016. Exploring Ethnic Inequalities in Admission to Russell Group Universities. *Sociology,* 50(2), pp. 247-267.

Bolton, P., 2021. *Higher education student numbers.* London: House of Commons (Library), (7857).

Bowes, L., Evans, J., Nathwani, T., Birkin, G., Boyd, A., Holmes, C., Thomas, L. and Jones, S., 2015. *Understanding progression into higher education for disadvantaged and under-represented groups.* London: BIS (Department for Business Innovation and Skills).

Bracke, P., Croxson, K., Leary, J. and Wood, J., 2021. Covid-19 and the UK's BAME communities – an economic perspective. *Financial Conduct Authority* [Online]. Available from: https://www.fca.org.uk/insight/covid-19-and-uk-bame-communities-economic-perspective [Accessed 20 February 2021].

Britton, J., Dearden, L., Shephard, N. and Vignoles, A., 2016. How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background. *IDEAS Working Paper Series from RePEc*.

Buis, M.L., 2010. Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal,* 10(2), pp. 305-308.

Callender, C., 2003. *Attitudes to Debt: School Leavers and Further Education Students' Attitudes to Debt and Their Impact on Participation in Higher Education: Ca Report for Universities UK and HEFCE.* London: Universities UK.

Callender, C. and Jackson, J., 2005. Does the Fear of Debt Deter Students from Higher Education? *Journal of Social Policy,* 34, pp. 509-540.

Cheng, S. and Long, J.S., 2007. Testing for IIA in the Multinomial Logit Model. *Sociological Methods & Research,* 35(4), pp. 583-600.

Christie, H., 2007. Higher Education and Spatial (Im)Mobility: Nontraditional Students and Living at Home. *Environment and Planning A,* 39(10), pp. 2445-2463.

Christie, H., Munro, M. and Fisher, T., 2004. Leaving university early: exploring the differences between continuing and non-continuing students. *Studies in Higher Education,* 29(5), pp. 617-636.

Collings, R., Swanson, V. and Watkins, R., 2014. The impact of peer mentoring on levels of student wellbeing, integration and retention: a controlled comparative evaluation of residential students in UK higher education. *Higher Education,* 68(6), pp. 927-942.

Connor, H., Tyers, C., Modood, T. and Hillage, J., 2004. *Why the difference? A closer look at higher education minority ethnic students and graduates.* Nottingham: DfES (Department for Education and Skills), (RR552).

Cramer, J.S. and Ridder, G., 1991. Pooling states in the multinomial logit model. *Journal of Econometrics,* 47(2-3), pp. 267-272.

Crawford, C., 2014a. *Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class.* Institute for Fiscal Studies, (Working Paper W14/31).

Crawford, C., 2014b. *The link between secondary school characteristics and university participation and outcomes.* London: Department for Education (Centre for Analysis of Youth Transitions), (DFE-RR353).

Crawford, C. and Greaves, E., 2015. *Socio-economic, ethnic and gender differences in HE participation.* London: BIS (Department for Business Innovation and Skills), (186).

Davies, R. and Elias, P., 2003. *Dropping Out: A Study of Early Leavers From Higher Education.* London: DfES (Department for Education and Skills), (RR386).

DfE, 2019. Universities must do more to tackle ethnic disparity. *Department for Education, The Cabinet Office* [Online]. Available from: *https://www.gov.uk/government/news/universities-must-do-more-to-tackle-ethnic-disparity* [Accessed 20 December 2020].

Dowd, B.E., Greene, W.H. and Norton, E.C., 2014. Computation of standard errors. *Health services research,* 49(2), p. 731.

Durkheim, E., 1951 [1897]. *Suicide: A Study in Sociology (translated to English by J.A. Spaulding & G. Simpson)* Glencoe, Illinois: Free Press.

Fielding, A., Belfield, C.R. and Thomas, H.R., 1998. The Consequences of Drop-outs on the Cost-effectiveness of 16-19 Colleges. *Oxford Review of Education,* 24(4), pp. 487-511.

Fry, T.R.L. and Harris, M.N., 1998. Testing for Independence of Irrelevant Alternatives: Some Empirical Results. *Sociological Methods and Research,* 26, pp. 401–423.

Ghignoni, E., 2017. Family background and university dropouts during the crisis: the case of Italy. *Higher Education,* 73(1), pp. 127-151.

Giani, M.S., Attewell, P. and Walling, D., 2020. The Value of an Incomplete Degree: Heterogeneity in the Labor Market Benefits of College Non-Completion. *The Journal of Higher Education,* 91(4), pp. 514-539.

Hartog, J., Pfann, G. and Ridder, G., 1989. (Non-)graduation and the earnings function: An inquiry on self-selection. *European Economic Review,* 33(7), pp. 1373-1395.

Hausman, J. and McFadden, D., 1984. Specification Tests for the Multinomial Logit Model. *Econometrica,* 52(5), pp. 1219-1240.

HEFCE, 2018. *Non-continuation rates: trends and profiles* [Online]. Bristol: HEFCE (Higher Education Funding Council for England). Available from: *https://webarchive.nationalarchives.gov.uk/20180405132816/http://www.hefce.ac.uk/media/HEFCE,2014/Content/Analysis/geography/Timeseries%20data.csv* [Accessed 20 December 2020].

Higher Education Commission, 2016. *From Bricks to Clicks - The Potential of Data and Analytics in Higher Education.* London: Policy Connect.

Hillman, N. and Robinson, N., 2016. *The underachievement of young men in higher education – and how to start tackling it.* Oxford: HEPI (Higher Education Policy Institute), (Report 84).

House of Commons, 2009. *Widening participation in higher education. Fourth Report of Session 2008–09.* London: House of Commons (Public Accounts Committee), (HC 226).

House of Commons, 2014. *Underachievement in education by white working class children: First report of session 2014-15.* London: House of Commons -The Education Committee, (HC 142).

Hovdhaugen, E. and Aamodt, P.O., 2009. Learning Environment: Relevant or Not to Students' Decision to Leave University? *Quality in Higher Education,* 15(2), pp. 177-189.

Hungerford, T. and Solon, G., 1987. Sheepskin Effects in the Returns to Education. *The Review of Economics and Statistics,* 69(1), pp. 175-177.

Jaeger, D.A. and Page, M.E., 1996. Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education. *The Review of Economics and Statistics,* 78(4), pp. 733-740.

Johnes, G. and McNabb, R., 2004. Never Give up on the Good Times: Student Attrition in the UK. *Oxford Bulletin of Economics and Statistics,* 66(1), pp. 23-47.

Johnes, J., 1990. Determinants of student wastage in higher education. *Studies in Higher Education,* 15(1), pp. 87-99.

Johnes, J., 1997. Inter-university variations in undergraduate non-completion rates: A statistical analysis by subject of study. *Journal of Applied Statistics,* 24(3), pp. 343-362.

Khambhaita, P. and Bhopal, K., 2015. Home or away? The significance of ethnicity, class and attainment in the housing choices of female university students. *Race Ethnicity and Education,* 18(4), pp. 535-566.

Lally, C., 2020. Impact of COVID-19 on different ethnic minority groups. *UK Parliament (Parliamentary Office of Science and Technology)* [Online]. Available from: https://post.parliament.uk/impact-of-covid-19-on-different-ethnic-minority-groups/ [Accessed 20 February 2021].

Larsen, M.R., Sommersel, H.B. and Larsen, M.S., 2013. *Evidence on dropout phenomena at universities.* Copenhagen: Danish Clearinghouse for Educational Research.

Layard, R. and Psacharopoulos, G., 1974. The Screening Hypothesis and the Returns to Education. *Journal of Political Economy,* 82(5), pp. 985-998.

Lee, C. and Buckthorpe, S., 2008. Robust Performance Indicators for Non-completion in Higher Education. *Quality in Higher Education,* 14(1), pp. 67-77.

Lee, L.-F., 1982. Specification error in multinomial logit models: Analysis of the omitted variable bias. *Journal of Econometrics,* 20(2), pp. 197-209.

Long, J.S. and Freese, J., 2014. *Regression models for categorical dependent variables using Stata.* 3rd ed. Texas: Stata Press.

Long, J.S. and Mustillo, S.A., 2018. Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes. *Sociological Methods & Research*, pp. 1-37.

Longden, B., 2012. “Bearing Down” on Student Non-Completion: Implications and Consequences for English Higher Education. *Journal of College Student Retention: Research, Theory & Practice,* 14(1), pp. 117-147.

Manski, C.F., 1989. Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review,* 8(4), pp. 305-312.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, ed. *Frontiers in Econometrics.* New York: Academic Press, pp. 105-142.

McGivney, V., 1996. *Staying or Leaving the Course: Non-Completion and Retention of Mature Students in Further and Higher Education.* National Inst. of Adult Continuing Education.

Mize, T.D., 2019. Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science,* 6, pp. 81-117.

Mood, C., 2009. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review,* 26(1), pp. 67-82.

National Audit Office, 2007. *Staying the Course: the retention of students in higher education.* London: The Stationery Office, (HC 616).

OECD, 2015. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence.* Organisation for Economic Co-operation and Development (OECD) Publishing - Programme for International Student Assessment (PISA).

OfS, 2018. *Topic briefing: Black and minority ethnic (BME) students.* OfS (Office for Students). Available from: https://www.officeforstudents.org.uk/media/145556db-8183-40b8-b7af-741bf2b55d79/topic-briefing_bme-students.pdf [Accessed 20 December 2020].

OfS, 2020. *TEF outcomes* [Online]. OfS (Office for Students),. Available from: https://www.officeforstudents.org.uk/advice-and-guidance/teaching/tef-outcomes/#/tefoutcomes/ [Accessed 15 June 2020].

ONS, 2021. Labour market overview, UK: January 2021. Estimates of employment, unemployment, economic inactivity and other employment-related statistics for the UK. . *Office for National Statistics* [Online]. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/january2021 [Accessed 20 February 2021].

Petrie, K. and Keohane, N., 2019. *Building on success - Increasing higher education retention in London.* London: SMF (Social Market Foundation). Available from: https://www.london.gov.uk/sites/default/files/building-on-success-increasing-student-retention-london-higher-education-smf.pdf [Accessed 20 December 2020].

Pforr, K., 2014. femlogit-Implementation of the multinomial logit model with fixed effects. *Stata Journal,* 14(4), pp. 847-862.

Quinn, J., 2013. *Drop-Out and Completion in Higher Education in Europe among Students from Under-Represented Groups.* Brussels: European Commission.

Rainey, C., 2016. Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction. *Political Science Research and Methods,* 4(3), pp. 621-639.

Ridley, A., 2007. Approaches to learning, age, ethnicity and assessment. Implications for widening participation. *Psychology Teaching Review,* 13(1).

Riley, J.G., 1979. Testing the Educational Screening Hypothesis. *Journal of Political Economy,* 87(5, Part 2), pp. S227-S252.

Rodgers, T., 2013. Should high non-completion rates amongst ethnic minority students be seen as an ethnicity issue? Evidence from a case study of a student cohort from a British University. *Higher Education,* 66(5), pp. 535-550.

Rose-Adams, J., 2013. Leaving university early: Exploring relationships between institution type and student withdrawal and implications for social mobility. *Widening Participation & Lifelong Learning,* 15(2), pp. 96-113.

Russell Group, 2019. *Supporting under-represented students to access and succeed in higher education.* Available from: https://russellgroup.ac.uk/media/5786/supporting-under-represented-students-to-access-and-succeed-in-he-july-2019.pdf [Accessed 20 December 2020].

Singh, R., 1990. Ethnic Minority Experience in Higher Education. *Higher Education Quarterly,* 44(4), pp. 344-359.

Small, K.A. and Hsiao, C., 1985. Multinomial Logit Specification Tests. *International Economic Review,* 26(3), pp. 619-627.

SMF, 2017. *On course for success? Student retention at university.* London: SMF (Social Market Foundation). Available from: http://www.smf.co.uk/wp-content/uploads/2017/07/UPP-final-report.pdf [Accessed 20 December 2020].

Smith, J.P. and Naylor, R.A., 2002. Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 164(2), pp. 389-405.

Spady, W.G., 1970. Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange,* 1(1), pp. 64-85.

Spence, M., 1973. Job Market Signaling. *The Quarterly Journal of Economics,* 87(3), pp. 355-374.

Stiglitz, J.E., 1975. The Theory of "Screening," Education, and the Distribution of Income. *The American Economic Review,* 65(3), pp. 283-300.

Thomas, L., Hill, M., O'Mahony, J. and Yorke, M., 2017. *Supporting student success: strategies for institutional change. What Works? Student Retention & Success programme.* London: HEFCE - Higher Education Academy - Paul Hamlyn Foundation - Action on Access. Available from: https://www.phf.org.uk/publications/works-student-retention-success-full/ [Accessed 20 December 2020].

Thomas, L. and Hovdhaugen, E., 2014. Complexities and Challenges of Researching Student Completion and Non-completion of HE Programmes in Europe: a comparative analysis between England and Norway. *European Journal of Education,* 49(4), pp. 457-470.

Tinto, V., 1975. Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research,* 45, pp. 89-126.

Tinto, V., 1982. Limits of Theory and Practice in Student Attrition. *The Journal of Higher Education,* 53(6), pp. 687-700.

Tinto, V., 1988. Stages of Student Departure: Reflections on the Longitudinal Character of Student Leaving. *The Journal of Higher Education,* 59(4), pp. 438-455.

Tinto, V., 1993. *Leaving College: Rethinking the Causes and Cures of Student Attrition.* 2nd ed. Chicago: The University of Chicago Press.

UCAS, 2020. *2019 End of Cycle Report: Equality in England.* Cheltenham: UCAS (Universities and Colleges Admissions Service), (MD-5927).

UUK and NUS, 2019. *Black, Asian and minority ethnic student attainment at UK universities: #Closingthegap.* London: Universities UK (UUK) and National Union of Students (NUS). Available from: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2019/bame-student-attainment-uk-universities-closing-the-gap.pdf [Accessed 20 December 2020].

Vignoles, A. and Powdthavee, N., 2009. The Socioeconomic Gap in University Dropouts. *The B.E. Journal of Economic Analysis & Policy,* 9(1).

Vossensteyn, J.J., Kottmann, A., Jongbloed, B.W., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E. and Wollscheid, S., 2015. *Dropout and completion in higher education in Europe.* Luxembourg: European Commission.

Walker, I. and Zhu, Y., 2018. University selectivity and the relative returns to higher education: Evidence from the UK. *Labour Economics,* 53, pp. 230-249.

Webb, O.J. and Cotton, D.R.E., 2018. Early withdrawal from higher education: a focus on academic experiences. *Teaching in Higher Education,* 23(7), pp. 835-852.

Yorke, M., 1998a. Non-Completion of Full-Time and Sandwich Students in English Higher Education: Costs to the Public Purse, and Some Implications. *Higher Education,* 36(2), pp. 181-194.

Yorke, M., 1998b. Non-completion of Undergraduate Study: some implications for policy in higher education. *Journal of Higher Education Policy and Management,* 20(2), pp. 189-201.

Yorke, M. and Longden, B., 2008. *The first year experience of higher education in the UK.* York: Higher Education Academy.

Yorke, M. and Thomas, L., 2003. Improving the Retention of Students from Lower Socio-economic Groups. *Journal of Higher Education Policy and Management,* 25(1), pp. 63-74.

Appendix of Chapter 2

Table 2.A1. Description of the variables used in this paper

Variable	Values	Description – Notes
Age	17-53 years	Age of student on the 31st of August in the reported academic year.
Disability	Yes, No	A binary variable showing whether a student has reported a disability.
Distance travelled	0-987 kilometers	This variable shows the distance (in a straight line) between student's home postcode (as reported before entry) and the institution's main campus location.
Ethnic group	White, Black Caribbean, Black African, Other Black, Indian, Pakistani, Bangladeshi, Chinese, Other Asian, Mixed, Other ethnic group	Student's self-reported ethnicity. I adopt the recommended classification based on the 2011 National Statistics. The White category includes students from White British, Irish, Gypsy or Irish Traveller, and any other White backgrounds. Although the White category contains some ethnic minorities, the inconsistent way institutions record its sub-categories across different countries or regions of the UK does not allow me to distinguish White British from the rest of White students. However, in line with the practice adopted in other studies in the field, the "ethnic minority" term used throughout this paper comprises only non-White persons. "Other ethnic group" includes Arab minorities and any other group not mentioned in any of the rest categories. HESA does not provide clarifications about the "Other Black" category. This group likely includes students who identify as "Black European" or "Black North American".
Home fees eligible	Yes, No	A binary variable showing whether a student is entitled to pay "home" tuition fees.
Length of programme	<= 2 years, 2-3 years, 3-4 years, 4-20 years.	Represents the expected duration (in years) of the programme, from the start of the study to the course's end. I dropped from the analysis very few observations with an unknown duration or an expected programme length over 20 years.
Major source of tuition fees	No award/financial backing, UK LEA mandatory award, Provider waiver, UK central government and Local Authorities, Other	Represents the students' primary source of tuition fees. "UK LEA mandatory award" comprises cases where the "Student Loans Company" (or the "Student Awards Agency" for Scotland) covers either the entire amount of tuition fees or a part of them (and students paying the remaining share). "Other" category includes charities, research councils, UK industries or student's employer, international agencies, and other overseas foundations.
Male	Yes, No	A binary variable capturing male students.
Mode of study	Part-time, Full-time, Sandwich, Other mode of study	"Part-time" includes individuals who studied on courses with a duration of fewer than 24 weeks per academic year and evening students. "Sandwich" covers students who attended a thin or thick sandwich course with study or placement amounting to at least 21 hours/week for no less than 24 weeks/academic year.

Continued on next page

Table 2.A1. (continued)		
Non-White/White staff ratio	0-0.72	For each university and academic year, this variable represents the ratio of non-White academic staff (including atypical) to White academic staff. I collected these figures from the publicly available HESA staff records and merged them with my main datasets.
Parental education	Parents without HE qualifications, Parents with HE qualifications, Unknown/refused	This captures students who reported that at least one of their parents/guardians holds a higher education qualification.
Proportion of non-White peers	0-1	Within the group of first-degree students, I calculated this variable at the university, course, and academic year level.
Region of university	London, North East, North West, Yorkshire, East Midlands, West Midlands, East of England, South East, South West, Wales, Scotland, N. Ireland	The location of higher education institution based on the postcode of its main administration premises.
Relative tariff score	0.16-4.32	This variable is the fraction of a student's tariff score over the average tariff score of his/her peers. It measures the student's relative academic ability. Within the group of first-degree students, the average tariff score is calculated at the university, course, and academic year level and represents the average peers' quality/ability.
Socio-economic classification	Higher managerial & professional occupations, Lower managerial & professional occupations, Intermediate occupations, Small employers & own account workers, Lower supervisory & technical occupations, Semi-routine occupations, Routine occupations, Never worked & long-term unemployed, Unknown/Not classified	This represents the occupation of student's parent/guardian with the highest earnings.
Staff-student ratio	0.004-0.86	For each university and academic year, this variable represents the ratio of academic staff (including atypical) to the total university's students. I collected these figures from the publicly available HESA student and staff records and merged them with my primary datasets.
Subject of study	Medicine & dentistry, Subjects allied to medicine, Biological sciences, Veterinary science, Agriculture & related subjects, Physical sciences, Mathematical sciences, Computer science, Engineering & technology, Architecture, building & planning, Social studies, Law, Business & administrative studies, Mass communications & documentation, Languages, Historical & philosophical studies, Creative arts & design, Education, Combined subject	The subject area of the first-degree course based on the 19 principal codes of the "Joint Academic Coding System" (JACS) classification. "Combined" subjects include joint degrees in one or over one code of subject, irrespective of the percentage contribution of each subject area (e.g., History (60%) & Politics (40%), Economics (90%) & Mathematics (10%)).
Subject of study (grouped areas)	STEM, LEM, Other subject, Combined subject	The STEM category covers subjects related to Science, Technology, Engineering, and Mathematics, as well as Architects and Health subjects. The LEM category includes subjects in Law, Economics, and Management. The remaining subjects (except for the combined ones) are grouped in the "Other" category. This grouping is only presented in Table 2.A2 of the Appendix.

Continued on next page

Table 2.A1. (continued)		
Tariff score	5-1,991 points	This is an aggregated score from student's prior qualifications. During the application process, the "Universities and Colleges Admissions Service" (UCAS) computes each student's total tariff points based on his/her qualifications and then provides them to HESA. This variable approximates the student's prior educational ability. The tariff score of 99% of the students included in this study is less than 670 points, whereas only 1% of them have less than 80 points.
TEF award	Gold, Silver, Bronze, Unknown	This variable shows the award (Gold, Silver or Bronze) given to the UK higher education providers, based on the Teaching Excellence Assessment. The six TEF metrics are: "teaching"; "assessment & feedback"; "academic support"; "continuation"; "employment/further study"; "highly skilled employment/further study". This variable is a measure of universities' quality for their undergraduate provision. I obtained the TEF outcomes from the Office for Students, as updated in June 2020 (OfS, 2020). The universities' engagement in the TEF assessment is currently voluntary.
Term-time accommodation	Own residence, Parental/guardian home, Provider's property, Private-sector halls, Other rented, Other, Not in attendance, Unknown	This variable defines the place where a student lived during the reported academic year. "Other rented" category refers to temporary arrangements (e.g., yearly house share). "Not in attendance" category captures students who were not in attendance at the university during the reported academic year because of industrial placement or other reasons (e.g., language year abroad).
Type of school	Public school, Private school, Unknown school type	Indicates the type of the previous provider attended by a student before entering higher education. "Public school" includes state-funded schools and colleges.
Type of university	Russell Group universities, Other pre-1992 institutions, Post-1992 institutions, Specialist institutions	Russell Group universities include 24 prestigious institutions. The 1992 Further and HE Act and following legislation resulted in abolishing the so-called "binary divide" between the centrally funded universities and the locally funded polytechnics. As a consequence, over 40 former polytechnics were granted degree-award status after 1992. The "Post-1992" category includes these former polytechnics along with the institutions established after 1992. "Specialist" institutions cover a small fraction of the analysis sample and mainly refer to colleges/universities of agriculture, arts, music and drama, etc.
University's average tariff score	164.1-612.8 points	This variable represents the average tariff score of institutions within each academic year. It is a measure of institution selectivity.
University's income per student	3.7-83.9 thousand pounds (£ thousand)	This variable is calculated as the fraction of each institution's total income over the number of its students. It is a measure of university quality. I collected these figures from the publicly available HESA finance records and merged them with my main datasets.
Year of student on course	1st, 2nd, 3rd, 4th and over	This variable shows the student's year number since joining the specific subject of study. It may differ from the programme's year if the undergraduate switches to another course or resits the year.

Source: HESA

Table 2.A2. Mean characteristics: White versus non-White students

Variable	White	non-White	Difference
Reason for ending course			
Successful completion	0.898	0.861	0.037^{***}
Academic failure	0.035	0.079	-0.044^{***}
Other reason (grouped)	0.067	0.060	0.007^{***}
-Health reasons	0.005	0.003	0.002^{***}
-Death	0.000	0.000	-0.000^{**}
-Financial reasons	0.001	0.003	-0.001^{***}
-Other personal reasons	0.033	0.026	0.008^{***}
-Written off after lapse of time	0.005	0.006	-0.001^{***}
-Exclusion	0.002	0.006	-0.004^{***}
-Gone into employment	0.004	0.002	0.002^{***}
-Other	0.016	0.015	0.001^{***}
Subject of Study (grouped)			
STEM	0.356	0.386	-0.030^{***}
LEM	0.178	0.268	-0.090^{***}
Other subject	0.278	0.145	0.132^{***}
Combined	0.188	0.200	-0.012^{***}
University type			
Russell Group	0.276	0.213	0.063^{***}
Other Pre-1992	0.204	0.207	-0.003^{***}
Post-1992	0.501	0.569	-0.068^{***}
Specialist	0.020	0.011	0.008^{***}
Region of university			
London	0.077	0.330	-0.253^{***}
North East	0.059	0.018	0.041^{***}
North West	0.131	0.093	0.038^{***}
Yorkshire	0.114	0.079	0.036^{***}
East Midlands	0.085	0.100	-0.015^{***}
West Midlands	0.069	0.123	-0.053^{***}
East of England	0.048	0.075	-0.028^{***}
South East	0.126	0.101	0.024^{***}
South West	0.094	0.034	0.059^{***}
Wales	0.068	0.020	0.049^{***}
Scotland	0.097	0.025	0.072^{***}
N. Ireland	0.031	0.002	0.029^{***}
Mode of study			
Part-time	0.023	0.053	-0.030^{***}
Full-time	0.886	0.860	0.026^{***}
Sandwich	0.075	0.064	0.011^{***}
Other mode of study	0.017	0.024	-0.007^{***}
Student's characteristics			
Male	0.449	0.463	-0.013^{***}
Age	20.637	20.796	-0.159^{***}
Disability	0.116	0.075	0.040^{***}
Home fees eligible	0.992	0.982	0.010^{***}
Term-time accommodation			
Own residence	0.097	0.100	-0.004^{***}
Parental/guardian home	0.225	0.420	-0.196^{***}
Provider's property	0.097	0.085	0.013^{***}
Other/unknown accommodation	0.581	0.394	0.187^{***}

Continued on next page

Table 2.A2. (continued)			
Socio-economic background (parental occupation)			
Higher managerial/professional	0.220	0.136	0.084***
Lower managerial/professional	0.261	0.203	0.059***
Intermediate	0.111	0.087	0.023***
Small employers	0.061	0.083	-0.022***
Technical	0.045	0.024	0.021***
Semi-routine	0.088	0.144	-0.056***
Routine	0.044	0.058	-0.014***
Long-term unemployed	0.001	0.004	-0.003***
Unknown occupation	0.169	0.262	-0.093***
Parental education			
Parents with HE qualifications	0.437	0.360	0.078***
Parents without HE qualifications	0.327	0.369	-0.042***
Unknown parental education	0.236	0.272	-0.036***
Pre-entry characteristics			
Tariff Score	354.296	315.939	38.357***
Private school	0.113	0.086	0.027***
Public school	0.848	0.859	-0.011***
Unknown school type	0.040	0.055	-0.015***
Distance travelled (km)	109.449	71.409	38.040***
Peer effects and other university quality measures			
Average tariff (peers)	350.252	328.957	21.296***
Proportion of non-White peers	0.152	0.427	-0.275***
University's average tariff score	347.968	328.872	19.096***
Staff-student ratio	0.115	0.112	0.003***
Non-White/White staff ratio	0.137	0.192	-0.055***
University's income per student (£ thousand)	12.919	12.536	0.383***
TEF-Gold	0.374	0.324	0.050***
TEF-Silver	0.461	0.518	-0.057***
TEF-Bronze	0.039	0.130	-0.091***
TEF-Unclassified	0.126	0.028	0.098***
Observations	**1,106,695**	**270,645**	**-**

Note: The equality of means between the two groups is examined using standard tests of proportions. The numbers of observations are rounded to the nearest multiple of 5, in line with data provider's disclosure control.

**p<0.1, **p<0.05, ***p<0.01.*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.A3. Proportion of students who dropped out for "other reasons" (voluntary dropout) by ethnic group and type of university

Ethnic group	**University type**				
	Russell Group	Other Pre-1992	Post-1992	Specialist	**Total**
White	0.04	0.05	0.09	0.10	**0.07**
Black Caribbean	0.03	0.05	0.09	0.13	**0.08**
Black African	0.04	0.04	0.08	0.12	**0.07**
Other Black	*	0.05	0.10	*	**0.08**
Indian	0.02	0.03	0.05	0.13	**0.04**
Pakistani	0.04	0.06	0.07	0.13	**0.07**
Bangladeshi	0.04	0.05	0.07	*	**0.06**
Chinese	0.02	0.03	0.05	*	**0.03**
Other Asian	0.03	0.04	0.07	*	**0.05**
Mixed	0.04	0.06	0.09	0.12	**0.07**
Other ethnic group	0.04	0.05	0.07	*	**0.06**
Unknown	0.03	0.06	0.09	*	**0.06**
Total	**0.04**	**0.05**	**0.08**	**0.10**	**0.07**

Note: "Other reasons" refer to voluntary dropout and include personal and financial reasons, departure to employment, exclusion, writing off after a period of inactivity, health reasons, and death.
** denotes cells with fewer than 23 students.*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.A4. Proportion of students who dropped out for "other reasons" (voluntary dropout) by ethnic group and subject of study

Subject of Study	Ethnic group														
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Medicine & dentistry	0.02	*	*	*	*	*	*	*	*	*	*	*	**0.01**	21,100	1.5%
Subjects allied to medicine	0.07	0.10	0.06	*	0.03	0.06	0.06	0.04	0.05	0.08	0.04	0.04	**0.07**	86,485	6.2%
Biological sciences	0.07	0.08	0.08	0.07	0.04	0.06	0.07	*	0.05	0.07	0.06	0.06	**0.07**	148,050	10.7%
Veterinary science	0.01	*	*	*	*	*	*	*	*	*	*	*	**0.01**	2,340	0.2%
Agriculture & related subjects	0.07	*	*	*	*	*	*	*	*	*	*	*	**0.07**	9,025	0.6%
Physical sciences	0.06	*	0.05	*	0.05	0.07	*	0.05	0.07	0.07	*	0.04	**0.06**	61,140	4.4%
Mathematical sciences	0.06	*	*	*	0.03	0.05	*	0.04	*	0.06	*	*	**0.05**	24,885	1.8%
Computer science	0.11	0.10	0.09	*	0.07	0.09	0.09	0.07	0.10	0.12	0.09	0.10	**0.10**	53,945	3.9%
Engineering & technology	0.07	0.08	0.07	*	0.06	0.08	0.10	0.04	0.07	0.07	0.08	0.07	**0.07**	66,940	4.8%
Architecture, building & planning	0.07	0.07	0.09	*	0.06	0.10	0.13	*	0.09	0.09	0.08	*	**0.07**	28,590	2.1%
Social studies	0.06	0.07	0.06	*	0.03	0.06	0.05	*	0.03	0.05	0.05	0.08	**0.06**	95,345	6.9%
Law	0.06	0.06	0.04	*	0.03	0.05	0.04	*	0.05	0.06	0.04	*	**0.05**	52,625	3.8%
Business & administrative studies	0.07	0.07	0.06	0.09	0.04	0.07	0.06	0.03	0.05	0.08	0.06	0.08	**0.07**	124,465	9.0%
Mass communications & documentation	0.08	0.08	0.08	*	0.06	0.12	*	*	*	0.10	0.10	*	**0.08**	35,630	2.6%
Languages	0.05	*	0.05	*	0.04	0.05	*	*	*	0.06	*	0.04	**0.05**	58,300	4.2%
Historical & philosophical studies	0.05	*	0.06	*	0.04	0.07	*	*	*	0.05	*	0.04	**0.05**	54,810	3.9%
Creative arts & design	0.08	0.10	0.10	0.12	0.08	0.11	0.09	0.05	0.08	0.09	0.08	0.08	**0.08**	150,645	10.8%
Education	0.07	0.13	0.15	*	0.05	0.08	0.09	*	*	0.10	*	0.14	**0.07**	50,080	3.6%
Combined	0.06	0.07	0.06	0.09	0.03	0.06	0.06	0.03	0.05	0.07	0.06	0.06	**0.06**	265,520	19.1%
Total	**0.07**	**0.08**	**0.07**	**0.08**	**0.04**	**0.07**	**0.06**	**0.03**	**0.05**	**0.07**	**0.06**	**0.06**	**0.07**	**1,389,920**	**100%**

Note: "Other reasons" refer to voluntary dropout and include personal and financial reasons, departure to employment, exclusion, writing off after a period of inactivity, health reasons, and death.
** denotes cells with fewer than 23 students. The total number of students (n) for each subject of study is rounded to the nearest multiple of 5, in line with data provider's disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

Table 2.A5. Proportion of students who dropped out for “other reasons” (voluntary dropout) by ethnic group and socio-economic background (parental occupation)

Socio-economic classification	Ethnic group														
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Higher managerial/professional	0.05	0.06	0.05	*	0.03	0.04	*	0.03	0.03	0.05	0.04	0.04	**0.05**	282,850	20.4%
Lower managerial/professional	0.06	0.07	0.06	0.07	0.04	0.06	0.05	0.03	0.05	0.07	0.05	0.04	**0.06**	346,510	24.9%
Intermediate	0.07	0.07	0.05	0.07	0.03	0.06	0.05	0.03	0.04	0.07	0.06	0.04	**0.06**	146,845	10.6%
Small employers/Own account workers	0.07	0.07	0.08	*	0.04	0.07	0.06	0.04	0.07	0.08	0.06	0.05	**0.07**	89,980	6.5%
Technical/lower supervisory	0.07	0.09	0.07	*	0.04	0.06	*	*	*	0.08	0.07	*	**0.07**	56,450	4.1%
Semi-routine	0.08	0.08	0.06	0.09	0.05	0.07	0.06	0.03	0.07	0.09	0.07	0.05	**0.08**	136,760	9.8%
Routine	0.09	0.10	0.08	*	0.05	0.07	0.07	*	0.07	0.10	0.08	0.09	**0.09**	64,270	4.6%
Long-term unemployed/Never worked	0.14	*	0.19	*	*	0.12	*	*	*	*	*	*	**0.13**	2,710	0.2%
Unknown	0.07	0.08	0.08	0.10	0.05	0.07	0.07	0.04	0.06	0.08	0.07	0.09	**0.07**	263,545	19.0%
Total	**0.07**	**0.08**	**0.07**	**0.08**	**0.04**	**0.07**	**0.06**	**0.03**	**0.05**	**0.07**	**0.06**	**0.06**	0.07	**1,389,920**	**100%**

Note: “Other reasons” refer to voluntary dropout and include personal and financial reasons, departure to employment, exclusion, writing off after a period of inactivity, health reasons, and death.
** denotes cells with fewer than 23 students. The total number of students (n) for each level of socio-economic classification is rounded to the nearest multiple of 5, in line with data provider’s disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author’s own calculations

Table 2.A6. A Wald test for combining alternative outcomes of the dependent variable

Combination	Chi-square	Degrees of freedom	p-value
Successful completion & Academic failure	1.32e+05	224	0.000
Successful completion & Voluntary dropout	1.72e+05	224	0.000
Academic failure & Voluntary dropout	20,305.14	224	0.000

Note: I performed the Wald tests shown in the table using the Stata mlogtest command (developed by Long and Freese (2014)) after running the multinomial logistic regression (MLR).
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.A7. Tests of independence of irrelevant alternatives (IIA) assumption

Hausman-McFadden (HM) test			
Excluded alternative	**Chi-square**	**Degrees of freedom**	**p-value**
Successful completion	-1073.99	211	.
Academic failure	24.102	17	0.117
Voluntary dropout (other reason)	0.000	1	1.000
Small-Hsiao (SH) test			
Excluded alternative	**Chi-square**	**Degrees of freedom**	**p-value**
Successful completion	349.648	224	0.000
Academic failure	536.762	225	0.000
Voluntary dropout (other reason)	713.442	225	0.000

Note: I performed the IIA tests shown in the table using the Stata mlogtest command (developed by Long and Freese (2014)) after running the multinomial logistic regression (MLR). For these tests' needs only, I did not use the robust standard errors (as reported in the main results section). Instead, I used the default variance estimator, which is based on the observed information matrix.
Both tests' null hypothesis is that the odds of outcome m versus the outcome n are independent of other alternatives, thus satisfying the IIA assumption. The negative value of the chi-square is usual when running the HM test. Hausman and McFadden (1984) claimed that negative values indicate that the IIA property holds, while Long and Freese (2014) warn that these values are problematic signifying that the estimated model does not fulfil the asymptotic properties.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.A8. Robustness analysis - Multinomial logistic regression: Average marginal effects (AMEs)

Dependent variable: Reason for ending course						
Variable	**Successful completion**		**Academic failure**		**Other reason**	
	AMEs	**SE**	**AMEs**	**SE**	**AMEs**	**SE**
Ethnic group						
White	+	+	+	+	+	+
Black Caribbean	-0.011***	(0.002)	0.025***	(0.002)	-0.014***	(0.002)
Black African	-0.012***	(0.001)	0.032***	(0.001)	-0.020***	(0.001)
Other Black	-0.011**	(0.004)	0.025***	(0.004)	-0.014***	(0.004)
Indian	0.001	(0.001)	0.012***	(0.001)	-0.014***	(0.001)
Pakistani	-0.008***	(0.001)	0.018***	(0.001)	-0.010***	(0.001)
Bangladeshi	-0.008***	(0.002)	0.021***	(0.003)	-0.013***	(0.002)
Chinese	0.001	(0.002)	0.011***	(0.002)	-0.012***	(0.002)
Other Asian	-0.006***	(0.001)	0.020***	(0.002)	-0.014***	(0.001)
Mixed	-0.008***	(0.001)	0.013***	(0.001)	-0.005***	(0.001)
Other ethnic group	-0.008***	(0.002)	0.023***	(0.002)	-0.014***	(0.002)
Observations			1,137,917			
Pseudo R^2			0.566			

Note: The marginal effects shown in the table are derived from the multinomial logistic regression (post-estimates). AMEs sum up to zero across all three outcomes.
SE: The standard errors in parentheses are estimated based on the delta method. In the multinomial logistic regression, robust standard errors are used (that is, the variance estimator is robust to certain misspecification forms).
+Reference category.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
*The model specification of this robustness analysis additionally includes the interaction term "Ethnicity*Subject of study". Unlike the sample of the regression presented in Table 2.5 of the main text, students who left university because of "health reasons" and "death" are excluded from the present models. The "Other reason" outcome refers to voluntary dropout and now includes: "financial reasons", "other personal reasons", "written off after lapse of time", "exclusion", "gone into employment", and "other".*
The cases with unknown ethnicity (<1% of the initial sample) are dropped from the regression analysis. For the dummy variables with a significant proportion of missing (unknown) values (>5%), I have included an additional category ("unknown").
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 2.A9. Robustness analysis - Interaction effects on the probability of academic failure (second difference approach)

Ethnic group	**Ethnicity*gender**	**Ethnicity*university type**			**Ethnicity*socioeconomic background**		
	Women – Men	**Russell group – Other pre-1992**	**Russell group – Post-1992**	**Other pre-1992 – Post-1992**	**Higher managerial – Small employers**	**Higher managerial – Routine**	**Small employers – Routine**
Black Caribbean	-0.014^{***}	-0.011	-0.015	-0.004	0.006	0.027^{**}	0.021
Black African	-0.010^{***}	-0.031^{***}	-0.031^{***}	-0.001	-0.007	-0.007	0.001
Other Black	-0.019^{*}	-0.094^{***}	-0.088^{***}	0.006	0.075^{**}	-0.012	-0.087^{**}
Indian	0.004^{***}	-0.005^{*}	-0.002	0.003	0.003	0.010^{***}	0.007^{*}
Pakistani	-0.011^{***}	-0.010^{*}	0.003	0.013^{***}	0.007	0.013^{**}	0.006
Bangladeshi	-0.009^{**}	-0.020^{***}	-0.023^{***}	-0.003	0.026^{*}	0.034^{**}	0.008
Chinese	0.004	0.005	0.008^{**}	0.003	0.016^{**}	0.016^{*}	0.000
Other Asian	0.001	-0.017^{***}	-0.011^{**}	0.006	-0.009	-0.009	0.000
Mixed	-0.009^{***}	-0.011^{***}	-0.018^{***}	-0.007	0.001	0.000	-0.001
Other ethnic group	-0.004	-0.017^{**}	-0.023^{***}	-0.006	-0.014	-0.005	0.008

Note: The columns show the difference in the ethnic gaps in the likelihood of academic failure between genders, type of institutions and socio-economic background (i.e., "second differences"). The results are derived from the original multinomial logistic regression (post-estimates), based on the same control variables and interaction terms as in Table 2.5. Specialist institutions are not presented in the table as they cover a small proportion of students (1.4%). The ethnic gap is defined as the difference in the average probability of dropping out due to academic failure between each ethnic minority and White students (reference category).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Statement of Authorship

<table>
<tr><th colspan="4">This declaration concerns the article entitled:</th></tr>
<tr><td colspan="4">Ethnic disparities in higher education attainment in the UK</td></tr>
<tr><td colspan="4">Publication status (tick one)
Draft manuscript ✓   Submitted ☐   In review ☐   Accepted ☐   Published ☐</td></tr>
<tr><td>Publication details (reference)</td><td colspan="3"></td></tr>
<tr><td colspan="4">Copyright status (tick the appropriate statement)
I hold the copyright for this material ✓   Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐</td></tr>
<tr><td>Candidate's contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="3">The candidate contributed to / considerably contributed to / predominantly executed the…

Formulation of ideas: 95%

Design of methodology: 100%

Experimental work: N/A

Presentation of data in journal format:100%</td></tr>
<tr><td>Statement from Candidate</td><td colspan="3">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td>Konstantinos Kollydas</td><td>Date</td><td>24/04/2021</td></tr>
</table>

3. Ethnic disparities in higher education attainment in the UK

Abstract

The UK Government has recently intensified its efforts to tackle ethnic disparities in students' academic performance. In light of these interventions, this study draws on large-scale data of over one million UK university graduates for the academic years 2010/11–2014/15 to investigate the extent of discrepancies in higher education attainment between ethnic minority students and their White counterparts. This work provides strong evidence that ethnic gaps in the probability of obtaining a "good" class of degree stubbornly persist, even after controlling for a plethora of personal socio-demographic traits, institutional factors, subject of study, prior attainment, and other pre-entry characteristics. White students are more likely to graduate with first-class or upper second-class honours (76%) than all ethnic minority groups, particularly with respect to Black minorities (61%-64%). Unlike studies to date, this paper uses a detailed ethnicity classification and explores interactive relationships to reveal heterogeneous effects of ethnicity on the likelihood of achieving a good degree, according to students' gender, social class, type of university attended, and previous educational ability. This study's findings should be of interest for national policymakers and universities to design and implement targeted policies that would improve the underperformance of ethnic minorities in the UK higher education.

Keywords: ethnic minorities, UK higher education, academic performance, ethnic attainment gaps, HESA

JEL classification: I24, J15

3.1 Introduction

A stated goal of the Government's national strategy on participation and success in higher education (HE) is to create a framework that guarantees equality of opportunity in enrolling and succeeding in the UK universities, regardless of a person's socio-economic background, ethnicity, age, gender, or disability status (BIS, 2014). The Government has recently launched measures to tackle academic performance disparities, mandating universities to publish attainment information broken down by student characteristics, such as ethnicity and socio-economic background (DfE, 2019b). In the context of widening participation, many higher education institutions have intensified their efforts to alleviate inequalities in attendance and academic attainment and to facilitate the success of disadvantaged students in their studies by providing them with the necessary support.

The disadvantaged background can also be explored along the lines of ethnicity, as ethnic minorities have higher poverty rates than the average population (Platt, 2007). It would be ethically questionable to foster the notion of widening participation among ethnic minorities without ensuring their equitable academic performance. This becomes of greater importance when considering the relevant literature on the economics of education, which suggests that performance in university is highly correlated with the labour market outcomes and career opportunities later in life. Specifically, students achieving a higher class of bachelor's degree are more likely to be employed in high-paid industries and have significantly higher salaries relative to graduates attaining a lower class of degree (Naylor, Smith and Telhaj, 2016; Feng and Graetz, 2017).

This paper is partly motivated by previous research, which has looked at the impact of ethnicity on academic achievement in UK higher education (for an updated literature review in this field, see Richardson, Mittelmeier and Rienties, 2020). However, most of the literature to date is based on relatively old data with limited information about some key factors that influence degree attainment, such as prior educational achievement and students' socio-economic background (e.g., Broecke and Nicholls, 2007). Furthermore, these analyses rarely examine the institutional-level variation in HE attainment, and their approaches do not consider the interactive relationship between specific variables when estimating the ethnic differences in the likelihood of obtaining a "good" class of degree. Quantifying the interdependence of the relevant characteristics that influence academic performance is a key element in evaluating the higher education experience of various ethnic groups.

This paper aims to fill this gap in the literature by examining the extent of ethnic differences in the probability of achieving a first-class or upper-second class of degree (that is, a “good” bachelor’s degree) in the UK for the academic years 2010/11–2014/15. Firstly, I descriptively investigate the primary determinants of the likelihood of gaining a good degree in relation to students’ socio-demographic and higher education characteristics. Secondly, I estimate the impact of ethnicity on the outcome in question and, for each ethnic group separately, I present the predicted probabilities of getting a good degree conditional on a rich set of characteristics (such as age, gender, disability, type of university, area of the subject of study, mode of study, type of term-time accommodation, socio-economic classification, parental education, pre-entry qualifications). Thirdly, I explore the heterogeneous effects of ethnicity on the likelihood of being awarded a good degree according to gender, the type of institution attended, and the students’ socio-economic background.

It has been well-documented for the last twenty-five years that ethnic minority students are less likely to graduate with a good degree compared to their White counterparts in the UK (Connor et al., 1996; Owen et al., 2000; Leslie, 2005; Broecke and Nicholls, 2007; Fielding et al., 2008; Richardson, 2008; HEFCE, 2018). For example, Richardson (2008) examined how the odds ratios of securing a first-class or upper second-class degree differed amongst four broader ethnic groups in 2004/05. He found that the odds ratios for students from Black, Asian, and Other ethnic backgrounds were 0.60, 0.71, and 0.77, respectively, compared to the reference White category. Although the coverage of graduates’ data on entry qualifications was limited at that time, the author showed that prior achievement accounted for nearly half of the attainment difference between White and non-White students. A more recent report carried out by the Higher Education Funding Council of England (HEFCE, 2018) indicates that the difference in the percentage of White and ethnic minority graduates holding a good degree has decreased only slightly between the 2013/14 and 2016/17 academic years.

Most of the works mentioned above neglect the significant role of intersectionality in explaining the “ethnic gaps” in higher education (that is, the difference in attainment between each ethnic minority group and their White peers). Indeed, they assume that the effect of ethnicity on the dependent variable is additive (i.e., independent of other variables), rather than operating interactively. In addition, because of data limitations, their analyses do not take into account specific variables (such as the distance between the student's home before entry and campus, and the region of HE provider) that may be correlated both with the ethnicity and the probability of being awarded a good degree, thus introducing additional bias in the estimates of interest. For instance, some ethnic minorities are less likely to relocate than White students, mainly because of cultural dissimilarities in their family's perspective towards gaining independence and underlying financial costs (Christie, 2007; Khambhaita and Bhopal, 2015). As a result, their new independent lifestyle likely affects their overall university experience differently than White students, with subsequent consequences for their academic performance. In this work, I eliminate the effect of such factors by explicitly including them in the regression models.

The contribution of this paper to the literature is twofold. First, I perform large-scale analysis of around one million observations using recent data from the Higher Education Statistics Agency (HESA) for academic years 2010/11 through to 2014/15 to estimate the likelihood of attaining a good degree for each ethnic group separately. The census nature of the data and the resultant large number of observations allow me to divide ethnicity into eleven categories. This detailed classification of ethnicity is a significant advantage of this study, as it enables me to consider inherent characteristics (such as motivation and cultural attitudes towards education), which are likely to differ among ethnic groups. Hence, this categorisation allows me to better explain the variation in the outcome of interest. A broader classification of ethnicity that is often adopted by other research studies might conceal ethnic dissimilarities. In addition, analysing the ethnic gaps at a more granular level will be of particular interest for national policymakers and universities to better understand the inequalities in academic attainment and, consequently, design and implement targeted policies that will focus on specific groups of students.

The second contribution is to incorporate and interpret interaction terms between ethnicity and student's gender, socio-economic background, type of university attended, and prior attainment into the logistic regression analysis. It may be reasonable and more realistic to expect that the effect of ethnicity on the probability of obtaining a good degree varies based on students' socio-economic class, their academic ability, and the level of university selectivity. However, interpreting the coefficients of the interaction between two independent variables is complicated for non-linear regressions (such as the binary logistic regression), which will be the preferred specification of models in the current analysis (Ai and Norton, 2003). Therefore, I apply recently developed approaches (Buis, 2010; Breen, Karlson and Holm, 2018; Long and Mustillo, 2018; Mize, 2019) to identify and present these interaction effects by testing whether the marginal effects are equal across different groups of students. To the best of my knowledge, this is the first study that considers interactive relationships in this way when analysing the achievement gaps between ethnic groups in UK higher education.

Throughout the paper, I focus on UK-domiciled young people (aged under 21 on entry), as they cover the vast majority (nearly 80%) of graduates with a classified first degree. Contrary to mature graduates, the information about specific important predictors of higher education attainment, such as the socio-economic classification and prior qualifications (Naylor, 2004; Richardson, 2015), is considerably more complete for young students in the HESA datasets. In addition, older students' academic performance may be correlated with some factors like family formation, level of earnings (for working students), experiences, and aspirations, which may vary across ethnic groups. However, the HESA data does not contain information about such characteristics and including mature students in the analysis would aggravate the bias of the estimated impact of ethnicity.

Because of data limitations, this study does not disentangle the effects of ethnicity from other unobserved factors of individuals (e.g., student's motivation and learning approaches, the interaction between students and faculty, discrimination in teaching and assessment), which almost certainly affect students' university experience and HE attainment (Ridley, 2007; Singh, 2011; Cotton et al., 2016). Therefore, this paper does not discern causality in the relationship between ethnicity and degree attainment. Richardson (2015) contends that ethnicity *per se* is not the effective characteristic determining academic performance but merely a proxy for other variables that are not observed in the administrative data and need to be identified.

The model results reveal that the average adjusted predicted probability of being awarded a good degree is significantly higher for White students (76.2%) than all ethnic minority groups, and the picture is particularly worrying for Black minorities. Specifically, the ethnic gaps range from 15.1 percentage points (for Black African students) to 4.3 percentage points (for the Mixed ethnic group). These ethnic gaps in the likelihood of achieving a good degree remain impressively robust even after performing specific sensitivity checks, such as including additional variables relating to peer effects and incorporating university and course fixed effects in the models.

Delving deeper in the results by investigating some interesting interaction effects, the ethnic gaps are larger amongst women than men across all Asian ethnic groups (Indian, Pakistani, Bangladeshi, and other Asian), except for Chinese. Moreover, the ethnic differences in attainment are statistically significant within all types of institutions. Notably, the ethnic gaps are smaller for Black students in Russell Group universities than the other pre-1992 and post-1992 institutions (with the difference between these types of institutions ranging from 1.6 to 5.8 percentage points), but larger for all Asian ethnic groups. Concerning the socio-economic background, which is proxied by parental occupation, the attainment differences are bigger for Asian students from high-class families than those from middle or lower-class backgrounds. For example, the ethnic gap for an average Chinese student whose father or mother is employed in a higher managerial/professional position is nearly twice (14.9 versus 8.0 percentage points) that of a Chinese student from a lower social class (routine occupations). In contrast, the underperformance of Black minorities (except for “other Black” students) is relatively stable across all levels of social class.

This paper aims to improve our understanding of the underlying dynamics associated with ethnic discrepancies in higher education attainment. By incorporating a large number of variables into the analysis and restricting the sample to students who lived in the UK before entering higher education, I have ruled out the impact of many factors, including but not limited to prior attainment (which may be viewed as a proxy for academic ability), socio-demographic traits, subject area of the first degree, other study characteristics, type and region of institution, and the fluency in the English language. As a corollary, any remaining gaps in HE attainment between ethnic groups should be attributable to unobserved characteristics, such as differences in learning styles and academic advising, discrimination, or cultural marginalisation, which all shape the overall university experience of students. This work provides evidence-based information about identifying policy interventions that would improve the academic performance and overall experience of ethnic minorities in the UK higher education. It could also deliver insights into detecting best practice among HE providers based on the types of institutions where attainment gaps are smaller than others.

The rest of this paper starts with discussing previous literature related to ethnic inequalities in higher education attainment (section 3.2). In section 3.3, I present descriptive evidence and describe the data and methodology used in this work. Section 3.4 provides the econometric model results, and section 3.5 concludes and outlines the policy implications of this study.

3.2 Previous literature

Ethnicity is a crucial element of social organisation. Many societies, including the UK, are structured based on a dominant ethnic group and several ethnic minorities, which usually differ in terms of language, culture, and religion. Previous studies have extensively examined the performance of ethnic minority students in primary and secondary education in the UK (Modood, 2003; Bradley and Taylor, 2004; Wilson, Burgess and Briggs, 2011; Meunier et al., 2013; Strand, 2015). For example, Strand (2015) indicates that all ethnic minority groups made significant progress in educational achievement (measured at the age of 16) during the period 1991-2006. As a result, the achievement gap between them and the White pupils narrowed substantially. For instance, in 1991, only 14% of Bangladeshi students attained "5+ GCSE A*-C", relative to 37% of White pupils. In 2006, the corresponding figures were 57% for Bangladeshi students against 58% for White pupils. Moreover, Strand shows that, over the 2004-2013 period, there has been a striking closing or disappearance of the achievement gap for almost all ethnic minority groups relative to White British students.

However, relatively few studies focused on discrepancies in higher education attainment amongst ethnic groups (Richardson, Mittelmeier and Rienties, 2020). In this section, I focus only on the works that study ethnic disparities in academic performance in UK universities.

In one of the earliest studies in the field, Connor et al. (1996) performed a survey to 1,177 alumni of four UK higher education institutions, and they found that only 39% of ethnic minority students had attained a good degree in 1993, versus 65% of the White graduates. Although their interview-based findings unveiled a mixed picture regarding the university experience of non-White students, the authors highlighted some cases where respondents from ethnic minorities reported racial discrimination and exclusion from typical sources of support, which likely undermined their academic performance. A research report by Owen et al. (2000), based on HESA data for the academic year 1998/99, descriptively showed that the proportion of non-White students graduating with first-class or upper second-class honours was notably lower (ranging from 25.1% for those self-identified as Black African to 45.9% for "Other" ethnic groups) than that of White graduates (53.1%). Although the percentage of students achieving good grades has remarkably increased since then, the trend of ethnic minorities' underperformance continued in later years (Connor et al., 2004; Leslie, 2005; Broecke and Nicholls, 2007; Fielding et al., 2008; HEFCE, 2010; Richardson, 2008, 2015; HEFCE, 2018).

Broecke and Nicholls (2007) relied on the 2004/05 graduate cohort to confirm that the academic under-attainment of ethnic minority students persisted even after allowing for differences in the impact of demographic and institutional factors in their econometric analysis. However, due to limited data coverage, their study did not control for some critical variables related to students' background, such as the school type attended prior to entry (private or public), and their socio-economic classification. Instead, the authors used the imperfect Index of Multiple Deprivation (IMD), which ranks neighbourhoods according to an average deprivation score, as a proxy for students' socio-economic background. A similar limitation was observed in the study conducted by Fielding et al. (2008), who performed a detailed analysis of this topic by exploring interaction effects between different variables. In the current analysis, I employ a wide range of background characteristics (including parental occupation and education, and type of prior institution) and acknowledge the role of interdependencies between ethnicity and student's socio-economic profile in exploring discrepancies in HE achievement.

Leslie (2005) claimed that ethnic minorities disproportionately choose specific subjects of study, where it is harder to achieve high grades, thus compromising their academic attainment. He also indicated that a large part of the disparity in performance between ethnic groups is attributable to the fact that ethnic minorities are not equally qualified to their White counterparts when entering higher education. While this argument is correct for most (but not for all) minorities (as alluded to in section 3.3), it could also reflect inequalities that some ethnic groups encounter in secondary schools (Shiner and Modood, 2002). As Richardson (2008, 2020) aptly points out, the “more means worse” statement is not a panacea for understanding and explaining academic achievement differences in higher education. Indeed, research suggests that HE participation rates saw more substantial increases for ethnic minority groups than their White British peers in the 2000s. As a result, all minority groups are now more likely to enrol at a university, even after controlling for individual and school background, as well as secondary school attainment (Crawford and Greaves, 2015). However, as I will demonstrate in section 3.4, controlling for prior educational qualifications or other background characteristics under no circumstances eliminates the ethnic gaps in the probability of being awarded a good degree.

Richardson (2015) performed a comprehensive literature review outlining the known factors that impact higher education underachievement of ethnic minorities and describing the uncharted territory in this research area. Summarising the findings from different studies over time, he shows that the underperformance of Black and Asian minorities is larger among mature students than in younger graduates, is more prominent for women relative to men, and this trend varies according to the subject area of degree. However, the author stresses that it is difficult to identify the confounding (unobserved) determinants that influence the ethnic gaps by solely using administrative data.

In the light of these disparities in academic attainment, Cotton et al. (2016) adopted a mixed-method approach (comprising online questionnaires and focus group interviews) and collected data from students and teaching staff across six different academic disciplines in the UK. The authors infer that differences in students’ motivation for participation in higher education and proficiency in English may be partly responsible for explaining the attainment gap among ethnic groups. In the present study, I reduce (if not eliminate) the effect of fluency in the English language on academic achievement by confining the sample of analysis to students who were schooled in the UK (i.e., “home” or UK-domiciled students).

3.3 Data and Methodology

3.3.1 Sample selection and variables

To address the research questions, I utilise pooled HESA data[12] for the academic years 2010/11–2014/15. The HESA datasets cover all students in higher education institutions in the UK and contain rich personal information on socio-demographic characteristics, family background, prior attainment, subject of degree, mode of study, and so on. The present analysis sample comprises 1.2 million first-degree graduates (qualifiers), who were UK-domiciled before commencing their course. Ethnicity data applies only to those graduates who lived in the UK before the beginning of their studies.

Following Richardson's (2008) approach, I have restricted my sample to graduates who received a classified honours degree. Including unclassified degrees in the analysis would prevent me from distinguishing graduates who attended an honours programme from those who completed a non-honours degree course. As Richardson mentions, combining graduates with general (ordinary) degrees with those achieving third-class honours would confound the qualification level with their achievement level. The reason is that some general (unclassified) degrees signify adequate performance in non-honours courses, whereas third-class and pass degrees imply unsatisfactory performance in honours courses[13].

To further improve the homogeneity of the sample, I have excluded all mature students (aged over 21 on entry), distance learners, and those studying at the "other undergraduate" level (such as "Higher National Diploma", "Certificate of Higher Education", and "Higher National Certificate"). As discussed earlier, the information about some critical academic performance predictors (including student's prior attainment and socio-economic background) is limited for mature graduates in the HESA data. For example, the missing cases for the *tariff score* variable (which summarises the previous attainment) correspond to over 80% of mature graduates, while the respective figure for socio-economic classification is

---

[12] *"HESA Student Record 2014/15; HESA Student Record 2013/14; HESA Student Record 2012/13; HESA Student Record 2011/12; HESA Student Record 2010/11. HESA declaration: Copyright Higher Education Statistics Agency Limited. Neither the Higher Education Statistics Agency Limited nor HESA Services Limited can accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by HESA Services."*

[13] *When including unclassified degrees in the present sample, they cover 4.7% of all degrees. Unclassified degrees are primarily observed in health-related disciplines (e.g., medicine and dentistry, and veterinary science) and Scottish universities.*

nearly 50%. In addition, the HESA data does not contain information about some characteristics that probably influence older students' performance to a greater extent than that of young students. These factors refer to family formation (especially for female students), income (for working students), experiences and aspirations, and are likely to differ between ethnic groups. Therefore, omitting these variables from the regression analysis would introduce further sources of bias in the estimated effect of ethnicity on the probability of getting a good degree if I included mature students in the sample. Postgraduate students are also excluded from the analysis, as HESA does not collect information about those students' attainment (that is, the data on the class of degree is restricted to the undergraduate qualifiers).

Table 3.A1 in the Appendix presents a detailed description of all variables recruited in the present analysis to predict the average probability of getting a good degree and estimate the ethnic gaps in attainment. I have chosen these factors based on the relevant literature in this field, as they are all known to influence the outcome of interest.

3.3.2 Descriptive evidence

Of the total 1,156,515 young students who graduated with an honours degree from the UK higher education institutions over the academic years 2010/11–2014/15, 80.7% regarded themselves as White. Table 3.1 presents the distribution of degree classes across ethnic groups. On average, all groups of ethnic minority students are less likely to achieve a good degree than their White counterparts. Specifically, 74.9% of White students were awarded a good degree during the considered period, while the corresponding share for Black students lay between 51%-53%. The students from Asian backgrounds perform better than Black minorities, but they still fall well below their White peers. The proportion of Chinese students graduating with first-class honours is comparable to that of White people (17.2% versus 19.9%). However, a higher percentage of the Chinese community graduated with lower degree classes (lower second and third/pass), which adversely affected their overall share of good degrees. In contrast to Black and Asian students, the ethnic gap in academic attainment is smaller for students from Mixed ethnic backgrounds, standing (on average) at 4.4 percentage points.

Table 3.2 presents the proportions of ethnic groups who graduated with a good degree from different types of institutions. On average, the students attending a Russell Group university are more likely to graduate with a good degree (84%), compared to their peers from the other pre-1992 (74%), post-1992 (65%), and specialist institutions (62%). However, the ethnic gaps remain across all types of universities. In section 3.4, I identify the magnitude and statistical significance of differences in academic attainment for each ethnic minority across different types of universities.

Similarly, White students outperform ethnic minorities across all subject areas of study (Table 3.3). In general, students in medicine & dentistry and veterinary science are more likely to secure a good degree. However, these results should be interpreted cautiously, as the proportion of students with classified degrees in most health-related disciplines is particularly small (see the corresponding note in Table 3.A1 of the Appendix). Languages and historical & philosophical studies see a high proportion of students attaining good grades (83%). In contrast, the corresponding figures in agricultural subjects (65%), Computer science (66%), Education (66%) and Architecture (67%) are lower than the overall average (72%).

As Table 3.4 shows, there is a clear correlation between the students' socio-economic profile (measured by parental occupation) and their likelihood of achieving good grades at university. 79% of graduates whose parents hold a higher managerial/professional position achieved a good degree in the period considered. The probability of attaining good grades declines as we move to lower segments of the socio-economic distribution. Specifically, students from middle-class backgrounds (e.g., small employers or technical occupations) have a 70% chance of graduating with a good degree. In comparison, the corresponding probability for students whose parents are occupied in low-skilled jobs (routine professions) or are long-term unemployed is only 65% and 57%, respectively. Across all socio-economic backgrounds, White students are more likely to perform better in higher education than their peers from ethnic minority backgrounds. Nevertheless, in section 3.4, I show that the ethnic gaps are not equal for all minorities at various social class levels.

Table 3.A2 in the Appendix synthesises the socio-demographic and academic profile of White and ethnic minority (non-White) students included in this analysis sample. On average, ethnic minority students are more likely to study LEM subjects (28.2%) than their White peers (18.3%). However, they are not equally represented in the elite Russell Group universities (22.3% versus

29.1%). Non-White students mainly choose institutions in London (32.7% compared to only 7.7% of White students). This is closely linked with their "average distance travelled" (72 km versus 112 km), as it is well-known that ethnic minority communities are largely concentrated in the capital of England. In addition, the parents/guardians of non-White students are less likely to hold a university qualification, and only 34.6% of them have managerial and professional jobs, compared to 49.1% of the White graduates' parents. The latter figures, combined with the smaller proportion of ethnic minorities attending private schools, underpin the argument that, on average, non-White students come from lower socio-economic backgrounds than their White counterparts. In the regression analysis (section 3.4), I consider these background characteristics to paint a comprehensive picture of ethnic inequalities in higher education attainment.

Most studies exploring academic performance in higher education use a measure of prior attainment as a predictor of students' ability. The UCAS tariff score is a suitable indicator of the previous achievement, as it summarises all student's qualifications at entry and the corresponding grades in a single metric. Figure 3.1 presents the average prior achievement of each ethnic group. The entry profile of Black students lags behind that of White students. This implies that performance gaps may develop at earlier stages of the educational cycle of these minorities, thus impairing their prospects for success in their later academic life. Interestingly, Chinese students show the highest levels of prior attainment (399 tariff points), followed by the White ethnic group (359 points) and the students from Mixed ethnic backgrounds (352 points).

Although the Chinese students, on average, outperform their White peers at the point of entry to HE, this advantage is not reflected in their subsequent higher education attainment. As I will show in section 3.4, the differences in HE attainment in favour of White students remain even after considering disparate choices of the subject of study and institution type. This suggests that although the prior educational ability is a strong predictor of HE performance, it should not be deemed as the sole cause of ethnic differences. Instead, other unobserved factors likely shape the ethnic minorities' under-attainment and are responsible for the remaining ethnic gaps. There is, consequently, a need for understanding and tackling the barriers driving the persistence of ethnic disparities in higher education performance.

Table 3.1. Distribution (%) of honours degrees by ethnic group and class of degree

Ethnic group	Degree Class						
	First	Upper second	Lower second	Third/ Pass	Good degree	*n*	%
White	19.9	55.1	21.7	3.4	74.9	933,880	80.7%
Black Caribbean	8.1	44.5	38.0	9.5	52.5	13,010	1.1%
Black African	8.1	43.2	38.5	10.2	51.2	31,310	2.7%
Other Black	8.0	43.5	37.2	11.3	51.5	2,235	0.2%
Indian	15.7	49.8	28.6	5.9	65.5	47,610	4.1%
Pakistani	11.3	45.0	35.5	8.2	56.3	30,130	2.6%
Bangladeshi	11.5	46.5	34.6	7.4	58.0	12,330	1.1%
Chinese	17.2	49.5	26.6	6.7	66.7	11,900	1.0%
Other Asian	13.8	47.1	31.7	7.5	60.9	15,970	1.4%
Mixed	16.8	53.7	24.8	4.7	70.5	37,705	3.3%
Other ethnic group	14.1	48.5	30.5	6.9	62.6	10,455	0.9%
Unknown	19.8	51.3	23.6	5.3	71.1	9,980	0.9%
Total	**18.7**	**53.7**	**23.5**	**4.1**	**72.4**	**1,156,515**	**100%**

Note: The total number of students (n) for each ethnic group is rounded to the nearest multiple of 5, in line with data provider's disclosure control.
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

Table 3.2. Proportion of good degree holders by ethnic group and type of university

Ethnic group	University type				
	Russell Group	Other Pre-1992	Post-1992	Specialist	**Total**
White	0.85	0.76	0.69	0.64	**0.75**
Black Caribbean	0.75	0.65	0.47	0.36	**0.53**
Black African	0.69	0.60	0.44	0.37	**0.51**
Other Black	0.74	0.58	0.47	0.35	**0.52**
Indian	0.78	0.71	0.57	0.40	**0.66**
Pakistani	0.70	0.62	0.50	0.47	**0.56**
Bangladeshi	0.70	0.63	0.54	0.36	**0.58**
Chinese	0.75	0.68	0.57	0.50	**0.67**
Other Asian	0.72	0.64	0.53	0.50	**0.61**
Mixed	0.83	0.74	0.62	0.60	**0.71**
Other ethnic group	0.78	0.68	0.54	0.48	**0.63**
Unknown	0.84	0.72	0.58	0.51	**0.71**
Total	**0.84**	**0.74**	**0.65**	**0.62**	**0.72**

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.3. Proportion of good degree holders by ethnic group and subject of study

Subject of Study	Ethnic group														
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Medicine & dentistry	0.92	*	0.87	*	0.90	0.83	0.80	0.91	0.89	0.90	0.88	0.98	**0.91**	3,645	0.3%
Subjects allied to medicine	0.73	0.57	0.57	0.53	0.72	0.63	0.66	0.72	0.64	0.72	0.68	0.67	**0.71**	71,095	6.1%
Biological sciences	0.72	0.52	0.50	0.47	0.65	0.58	0.67	0.71	0.60	0.68	0.60	0.63	**0.70**	125,675	10.9%
Veterinary science	0.87	*	*	*	*	*	*	*	*	*	*	*	**0.86**	365	0.0%
Agriculture & related subjects	0.65	*	*	*	0.66	*	*	*	*	0.68	*	*	**0.65**	7,630	0.7%
Physical sciences	0.73	0.52	0.48	*	0.65	0.51	0.58	0.64	0.61	0.69	0.62	0.70	**0.72**	53,100	4.6%
Mathematical sciences	0.74	0.73	0.56	*	0.66	0.65	0.67	0.64	0.59	0.67	0.60	0.78	**0.72**	21,445	1.9%
Computer science	0.70	0.46	0.43	0.55	0.59	0.51	0.55	0.60	0.55	0.64	0.57	0.60	**0.66**	38,105	3.3%
Engineering & technology	0.76	0.55	0.56	0.60	0.69	0.62	0.56	0.70	0.61	0.70	0.66	0.69	**0.73**	51,645	4.5%
Architecture, building & planning	0.70	0.41	0.43	*	0.51	0.49	0.46	0.62	0.51	0.67	0.53	0.62	**0.67**	23,745	2.1%
Social studies	0.75	0.52	0.52	0.50	0.68	0.53	0.56	0.73	0.66	0.72	0.67	0.74	**0.72**	83,130	7.2%
Law	0.75	0.54	0.54	0.52	0.63	0.52	0.56	0.75	0.61	0.68	0.64	0.66	**0.70**	45,825	4.0%
Business & administrative studies	0.74	0.55	0.50	0.52	0.63	0.56	0.59	0.62	0.59	0.68	0.59	0.62	**0.70**	103,610	9.0%
Mass communications & documentation	0.74	0.54	0.47	0.47	0.61	0.56	0.46	0.65	0.59	0.68	0.60	0.62	**0.71**	29,890	2.6%
Languages	0.84	0.65	0.66	0.71	0.74	0.61	0.67	0.81	0.71	0.81	0.71	0.87	**0.83**	53,065	4.6%
Historical & philosophical studies	0.84	0.66	0.66	*	0.81	0.66	0.62	0.87	0.76	0.83	0.80	0.85	**0.83**	50,150	4.3%
Creative arts & design	0.73	0.45	0.41	0.48	0.54	0.46	0.41	0.55	0.54	0.66	0.52	0.65	**0.71**	128,195	11.1%
Education	0.68	0.45	0.39	*	0.52	0.49	0.45	0.61	0.48	0.60	0.53	0.54	**0.66**	43,400	3.8%
Combined	0.77	0.55	0.52	0.53	0.66	0.57	0.59	0.67	0.63	0.73	0.64	0.76	**0.74**	222,800	19.3%
Total	**0.75**	**0.53**	**0.51**	**0.52**	**0.66**	**0.56**	**0.58**	**0.67**	**0.61**	**0.71**	**0.63**	**0.71**	**0.72**	**1,156,515**	**100%**

*Note: * denotes cells with fewer than 23 students. The total number of students (n) for each subject of study is rounded to the nearest multiple of 5, in line with data provider's disclosure control.*
Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

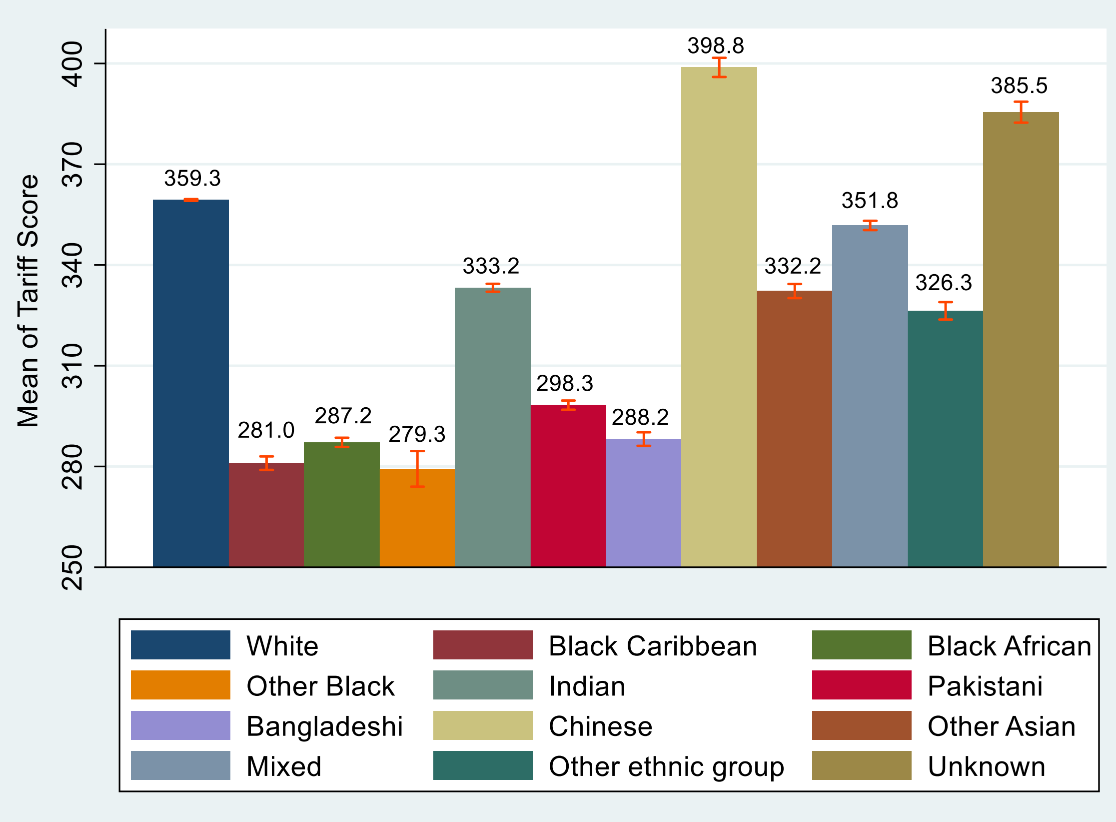
Table 3.4. Proportion of good degree holders by ethnic group and socio-economic background (parental occupation)

Socio-economic classification	**Ethnic group**														
	White	**Black Caribbean**	**Black African**	**Other Black**	**Indian**	**Pakistani**	**Bangladeshi**	**Chinese**	**Other Asian**	**Mixed**	**Other ethnic group**	**Unknown**	**Total**	**n**	**%**
Higher managerial/professional	0.80	0.59	0.59	0.63	0.72	0.64	0.66	0.73	0.69	0.77	0.73	0.81	**0.79**	242,725	21.0%
Lower managerial/professional	0.76	0.55	0.54	0.55	0.67	0.58	0.61	0.68	0.63	0.72	0.65	0.77	**0.74**	293,525	25.4%
Intermediate	0.75	0.54	0.52	0.55	0.67	0.63	0.61	0.68	0.66	0.71	0.66	0.74	**0.73**	123,985	10.7%
Small employers/Own account workers	0.73	0.55	0.53	0.48	0.63	0.55	0.61	0.68	0.61	0.69	0.63	0.69	**0.70**	74,130	6.4%
Technical/lower supervisory	0.72	0.52	0.54	0.52	0.63	0.55	0.64	0.64	0.60	0.67	0.62	0.71	**0.71**	47,045	4.1%
Semi-routine	0.70	0.50	0.49	0.48	0.64	0.57	0.59	0.68	0.57	0.65	0.62	0.67	**0.67**	110,185	9.5%
Routine	0.67	0.50	0.49	0.42	0.62	0.57	0.60	0.65	0.59	0.63	0.62	0.64	**0.65**	50,720	4.4%
Long-term unemployed/Never worked	0.61	*	0.52	*	0.49	0.50	0.47	*	0.62	0.62	0.59	*	**0.57**	1,840	0.2%
Unknown	0.72	0.48	0.47	0.47	0.62	0.53	0.54	0.61	0.56	0.67	0.55	0.64	**0.68**	212,360	18.4%
Total	**0.75**	**0.53**	**0.51**	**0.52**	**0.66**	**0.56**	**0.58**	**0.67**	**0.61**	**0.71**	**0.63**	**0.71**	**0.72**	**1,156,515**	**100%**

*Note: * denotes cells with fewer than 23 students. The total number of students (n) for each level of socio-economic classification is rounded to the nearest multiple of 5, in line with data provider's disclosure control.*

Source: HESA (pooled data for the academic years 2010/11–2014/15), author's own calculations

Figure 3.1. Average tariff score by ethnic group



Note: The error bars represent the 95% confidence interval of the mean tariff score of each ethnic group.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

3.3.3 Methodology

3.3.3.1 Logistic regression specification

The main interest of this study is to explain a qualitative event with a binary outcome. This means that the dependent variable (Y) can take only two values (typically coded as one and zero), which indicate whether a university student graduates with a good degree (first-class or upper second-class honours) or not. The probability of this event depends on several characteristics gathered in a vector $\mathbf{x}$, so that:

$$Prob(Y_i = 1) = F(\mathbf{x_i}, \boldsymbol{\beta}),$$

$$Prob(Y_i = 0) = 1 - F(\mathbf{x_i}, \boldsymbol{\beta}).$$

The set of parameters $\boldsymbol{\beta}$ reflects the impact of changes in $\mathbf{x}$ on the likelihood of the i–th individual obtaining a good degree. The challenge is to develop a precise mathematical form of the function $F(*)$. The linear probability model

(LPM) would be a simple method to estimate the likelihood of achieving a good degree. This method has some shortcomings, such as the fact that the fitted probabilities can take values less than zero or greater than one. This drawback of the LPM can sometimes result in unrealistic probability estimates (Mood, 2009). Nevertheless, this should not be a severe issue in the current context because of the large number of observations used in the analysis. More importantly, linear models assume that the effect of an un-interacted independent variable is constant across all its values. For example, the LPM implies that a change in a student's tariff score by a discrete amount (say, increasing from 50 tariff points to 100 points) would have the same impact on the probability of getting a good degree regardless of the tariff score's level (that is, the impact would be the same as if tariff points raised from 400 to 450). Thus, from this perspective, the LPM might exacerbate the functional form misspecification, as it is plausible that the relationship between a binary response variable and the covariates is non-linear.

The literature typically relies on the logistic function (for the logistic regression) and the cumulative normal distribution (for the probit regression) to estimate a non-linear relationship when the predicted variable is binary. Amemiya (1981) suggested that both forms provide satisfactory results, and neither has any significant advantage over the other. The expected value for a standard normal and a logistic random variable is zero, but the latter has a variance of $\frac{\pi^2}{3}$ (instead of 1), which leads to the form shown in equation (1) below. In addition, the distribution function in the logit model has slightly "flatter tails". However, both functions produce similar results in empirical works.

The present analysis adopts the binary logistic regression models to estimate the probability of an individual graduating with a good degree (π_i), given different values of the independent variables. The model form is:

$$\pi_i = Prob(Y_i = 1|X_i = x_i) = \frac{\exp(\mathbf{x'}_i\boldsymbol{\beta})}{1+ \exp(\mathbf{x'}_i\boldsymbol{\beta})} \qquad (1)$$

where Y_i is the binary response variable (which is equal to 1 if the student i graduates with a good degree), $\mathbf{x}_i$ is a set of explanatory variables (in a matrix notation), which can be discrete or continuous, and $\boldsymbol{\beta}$ is a vector of unknown parameters.

The vector $\mathbf{x}$ contains the primary variable of interest (ethnicity), all the covariates described in the previous section, and the interaction terms discussed in the following subsection. In particular, the logit models control for socio-demographic traits (gender, age, disability status, socio-economic background, and parental education); institutional and study characteristics (type and region of the university, subject of study, mode of study, programme's length); pre-entry factors (tariff score, type of school, distance travelled from student's home before entry to the institution); and other individual-level characteristics (term-time accommodation, home fees eligibility, source of tuition fees). I also include academic year fixed effects in vector $\mathbf{x}$ to capture the fact that, along with the higher education expansion and the increase in HE participation rates (DfE, 2019a), the proportion of students graduating with good grades continues to grow from cohort to cohort.

3.3.3.2 Interaction effects

Unlike linear regression, in non-linear models, the impact of a specific variable on the predicted probability is interactive in nature. The "compression" aspect of the *S*-shaped logistic function suggests that the effect of ethnicity on the outcome of interest is not constant across different levels of the curve (that is, across different values of the rest independent variables), even if no interaction terms are included in the model (Berry, Demeritt and Esarey, 2010). However, despite this inherent interdependent nature of logit models, neglecting to explicitly include a significant product term in the regression is likely to have severe implications when inferring interactive relationships (Rainey, 2016; Mize, 2019). Specifically, omitting a significant interaction term of interest does not allow the slopes of the curves of the logistic function to vary across different predictors' levels, which may, in turn, lead to wrong conclusions regarding the interactive relationship at issue.

In this paper, I consider heterogeneous effects by incorporating interaction terms into the analysis. This allows me to investigate whether ethnicity has significantly different impacts on the likelihood of achieving a good degree for students with specific characteristics, keeping all other regressors fixed. To avoid the negative consequences of model over-parameterisation, I concentrate on the interaction effects of higher interest in addressing the research questions. In this context, the vector $\mathbf{x}$ in the right-hand part of equation (1) also contains the following interaction terms: "*Ethnicity*Gender*", "*Ethnicity*University type*", "*Ethnicity*Socio-economic classification*", and "*Ethnicity*Tariff Score*". For example, the interaction effect between ethnicity and socio-economic classification tells by how much the impact of ethnicity on the dependent variable differs between graduates from diverse socio-economic backgrounds, in multiplicative terms. As I will show in section 3.4.3 (robustness checks), I also experimented with adding a few more interaction terms (such as "*Ethnicity*Subject of study*") in the models, but the results regarding ethnicity's effect on the probability of obtaining a good degree remained remarkably robust.

Interpreting interaction effects in non-linear regressions is not a straightforward process. Ai and Norton (2003) explain that in binary outcome models, the interaction effect should not be assessed only by focusing on the statistical significance, sign, and size of the relevant product term's coefficient (see also Buis, 2010). They illustrate that a consistent estimator for the interaction effect in non-linear models is derived from computing cross derivatives (or cross differences, with categorical regressors) of *Y*'s expected value, and the interaction effect is conditional on the explanatory variables. Hence, the marginal effects discussed below provide a solution for effectively presenting and interpreting interaction effects in non-linear specifications.

3.3.3.3 Marginal effects

Estimating marginal effects (MEs) is an effective way to summarise the impact of ethnicity on the dependent variable, as they are based on the predicted probabilities derived from the original logistic regression (post-model estimation). They represent a single and easily interpretable measure of each ethnic minority's effect on the likelihood of attaining a good degree, even when interaction terms are included in the model (Long and Freese, 2014). Also, this measure overcomes the well-known problem associated with identification (scaling) issues in binary outcome models (such as the logistic regression), which renders the traditional comparisons of coefficients unreliable (Amemiya, 1981; Allison, 1999). In the current analysis, I, therefore, follow the approach suggested by Long (see, for example, Long and Mustillo, 2018) and rely on the predicted probabilities (instead of the raw regression coefficients) to test the equality of ethnicity's effect across groups (e.g., across genders, types of institutions, and levels of socio-economic background).

Equation (2) shows how the average marginal effects (AMEs) are calculated for ethnic minorities relative to White graduates. For each ethnic minority group ($ethnic = k$), the AMEs represent the difference in the average predicted probability (π_i) of getting a good degree between the ethnic group k and the White reference category (that is, the ethnic gap), conditional on the observed values in the data of the rest independent variables. For example, the AMEs for Chinese graduates ($k =$ *Chinese*) are intuitively calculated as follows (Williams, 2012). Consider the first observation ($i = 1$) in the data as though he/she were Chinese (regardless of the actual ethnicity of that person). Keeping values of all other covariates as is, compute the probability π_1. Likewise, now treat the first observation assuming they were White and estimate the difference in these two probabilities (i.e., the first person's marginal effect). Similarly, repeat the process for all observations in the sample (n) and calculate the average MEs (AMEs).

$$AME_{ethnic_k} = \frac{1}{n}\sum_{i=1}^{n}(\pi_i\ (ethnic = k, \mathbf{x} = \mathbf{x}_i) - \pi_i\ (ethnic = White, \mathbf{x} = \mathbf{x}_i)) \quad (2)$$

To examine whether the AMEs for each ethnic group are equal across different levels of other variables (that is, to test the interactive effects), I use the "second differences" approach (Mize, 2019). For instance, let $\hat{\Delta}_{ethnic_k|women}$ and $\hat{\Delta}_{ethnic_k|men}$ be the AMEs of the ethnic minority group k for women and men, respectively. A second-difference test shows whether the ethnic gap in degree attainment for the ethnic group k differs in a statistically significant way between women and men. The denominator in equation (3) represents the estimated standard errors and covariance between the two marginal effects. Statistical software packages commonly use the *delta method* (Agresti, 2013; Dowd, Greene and Norton, 2014) to compute the variances of the partial effects. A Wald test is performed to examine the equality of the marginal effects.

$$z = \frac{\widehat{\Delta}_{ethnic_k|women} - \widehat{\Delta}_{ethnic_k|men}}{\sqrt{\widehat{\sigma}^2_{ethnic_k|women} + \widehat{\sigma}^2_{ethnic_k|men} - 2\widehat{\sigma}_{ethnic_k|women\,,\,ethnic_k|men}}} \quad (3)$$

Given that there are different ways to compute marginal effects depending on the situation (Long and Freese, 2014), I also present the marginal effects at the means (MEMs) in this study. Instead of relying on the observed values in the data, this post-estimation approach represents the ethnic gaps holding the other independent variables' values at their mean. Some reservations regarding the MEMs method are that the "average person" may not correspond to any existing values in the sample (for example, a graduate cannot be 56% female). Therefore, some authors (e.g., Cameron and Trivedi, 2005) recommend that it is best to choose AMEs over MEMs. As alluded to in section 3.4, the results between both approaches (AMEs and MEMs) are similar. In the results section, I also present the average adjusted predictions (AAPs) and the adjusted predictions at means (APMs) for each ethnic group. AAPs represent the conditional probability of achieving a good degree, keeping the rest regressors' values as is, while APMs are the corresponding estimated probabilities by holding the other explanatory variables' values at their mean.

3.3.3.4 Caveats

It is important to clarify that although the models include a comprehensive set of control variables that influence academic performance, this study will not necessarily reveal the causal effect of ethnicity on the probability of attaining a good degree because of the potentially endogenous nature of the main independent variable of interest. The problem of measuring causal effects is complicated by the fact that several unobserved factors may be correlated with both the key independent variable (i.e., ethnicity) and the outcome (probability of achieving a good degree). As a result, the coefficient estimates for each ethnic group could be biased. Examples of omitted variables include students' learning styles, cultural attitudes towards education, individual aspiration or self-motivation, and discrimination in teaching support and assessments. For instance, if self-motivation is, on average, higher for a specific ethnic group and it is also positively correlated with academic achievement, then not including this variable in the analysis would result in an upward bias in the estimate of this ethnic group's population parameter. Similarly, some ethnic minorities might use less efficient methods of studying (Ridley, 2007), which negatively affect their probability of achieving a good degree. Therefore, omitting the variable capturing the learning styles from the models may overestimate the size of ethnic gaps in academic attainment.

The extensive sample size used here provides a remedy for multicollinearity issues in the relationship between ethnicity and other predictors of academic attainment. However, some independent variables are likely on the “causal path” of ethnicity. For instance, if there is a causal effect of ethnicity on prior attainment, then estimating the impact of ethnic minority groups on the likelihood of getting a good degree would be problematic when the models include the *tariff score* variable, thus biasing the coefficients of ethnicity (i.e., the treatment effect). More specifically, if ethnicity lowers prior attainment, then incorporating *tariff score* in the regression specification translates into comparing ethnic minority students representing a higher slice of the ability distribution with less able White students. Consequently, the true ethnic gap in the likelihood of earning a good degree would be underestimated. Other examples of factors that might lie on the causal pathway between ethnicity and the outcome are the school type and the type of higher education institution, as ethnic minorities are less likely to attend private schools and prestigious universities than White students.

3.4 Results

This section presents the relationship between the probability of achieving a good degree and all the independent variables contained in the logistic regression. I mainly focus on ethnicity, which is the headline figure. Throughout this section, I interpret the results centring on the AMEs and the average adjusted predictions (AAPs). However, I also display the MEMs and the adjusted predictions at means (APMs) for comparison and completeness purposes. Moreover, I interpret the interactive effects[14] between the variable of interest (ethnicity) and gender, type of institution, socio-economic background, and prior attainment.

3.4.1 Average marginal effects and adjusted predictions

The results illustrate that there are significant ethnic gaps in the likelihood of graduating with good grades, even after controlling for all the factors described in the previous section. In sum, White students are more likely to achieve good grades at university than all ethnic minority groups (see Tables 3.5 and 3.6). These disparities in higher education attainment between ethnic minorities and

---

[14] *All interaction terms included in the logistic regression are statistically significant at the 1% level (based on the chi-square contrasts). However, as discussed in section 3.3, to evaluate and interpret the interaction effects in non-linear models, I focus on the average marginal effects and apply the “second difference” approach.*

White students are robust. Specifically, including a plethora of factors in the regression analysis only partly mitigates the raw differences in the average probability of gaining a good degree, as presented in the preceding descriptive analysis section. This reaffirms that ethnic disparities constitute a well-established, unfavourable situation in UK higher education, and it is crucial to identify the underlying mechanisms responsible for these inequalities.

Table 3.5 and Figures 3.2-3.3 reveal that the ethnic gaps in the probability of being awarded a good degree are more pronounced amongst Black students, ranging from 11.9 percentage points (for Black Caribbean students) to 15.1 percentage points (for the Black African minority). The underperformance in higher education is relatively smaller but still substantial for Asian communities, varying from 7 percentage points for Indian graduates to 11.1 points for Chinese people. In contrast, the picture is more encouraging for graduates from Mixed ethnic backgrounds, as the gap in their probability of achieving a good degree (compared to White students) stands at 4.3 percentage points, keeping all other variables fixed.

Unsurprisingly, students from high socio-economic backgrounds have a greater probability of graduating with good grades than others. At the same time, parental education is also positively associated with the likelihood of gaining first-class or upper second-class honours. On average, women show stronger academic performance (5.3 percentage points higher) than men. Furthermore, full-time students and those attending courses with an industry placement ("sandwich" courses) are, by far, more likely to perform well at university than part-time students (the difference stands at 31.3 and 37.9 percentage points, respectively). The latter figures are consistent with the findings of Richardson (2008) and other studies in the field.

The graduates of Russell Group universities are, on average, more likely to secure a good degree relative to the alumni of other types of institutions, while the probability of achieving good grades varies from subject to subject. For instance, Mathematics students have a significantly lower probability of being awarded a good degree than those from Social sciences (11.7 percentage points less, on average). Also, there is evidence that the so-called "London effect" observed in secondary education (Greaves, Macmillan and Sibieta, 2014) is followed through in higher education attainment. Specifically, people studying in London are more likely to fare well at university than all other UK regions (conditional on completing their degrees). As expected, prior attainment is positively linked with academic achievement. Moreover, the further from university the students lived before the commencement of their studies, the more likely they were to achieve good grades, keeping all else equal. The latter finding could be another reflection of socio-economic status, as students from prosperous families are more likely to study a long distance away from home than those from poorer families.

Students who attended independent (private) schools before entering higher education are, on average, less likely to achieve a good degree at university than equivalent state school graduates. This is in line with previous research findings (e.g., Crawford, 2014; Vidal Rodeiro and Zanini, 2015). These studies claim that one reason for the underperformance of private-school students might be that they have a lower motivation to get good grades at university, as they focus on spending more time in social activities and maintaining their parents' social class status. The second reason is related to the teaching they receive at school. Private schools attract more qualified teachers and invest more financial resources than state schools to support their students in attaining high grades at their examinations[15]. Hence, the second argument implies that because private school pupils are coached closely by their teachers at school, they find it harder to perform well when they enrol at university and work independently.

[15] *Indeed, the average* tariff score *in the analysis sample is significantly higher for private school students (407 points) than their peers in state schools (346 points). Apart from the teacher and school quality, this difference also likely reflects the higher socio-economic background and parental education of independent school students.*

Table 3.5. Logistic regression: Average marginal effects (AMEs) and Marginal effects at means (MEMs)

Dependent variable: Good degree

Variable	AMEs	MEMs	Variable	AMEs	MEMs
Ethnic group			***University type***		
White	+	+	Russell Group	+	+
Black Caribbean	-0.119*** (0.006)	-0.128*** (0.007)	Other pre-1992	-0.017*** (0.001)	-0.019*** (0.001)
Black African	-0.151*** (0.004)	-0.168*** (0.004)	Post-1992	-0.033*** (0.001)	-0.036*** (0.001)
Other Black	-0.139*** (0.014)	-0.154*** (0.017)	Specialist	-0.100*** (0.004)	-0.107*** (0.005)
Indian	-0.070*** (0.002)	-0.078*** (0.003)	***Subject of Study***		
Pakistani	-0.094*** (0.004)	-0.106*** (0.004)	Social studies	+	+
Bangladeshi	-0.078*** (0.007)	-0.090*** (0.008)	Medicine & dentistry	-0.044*** (0.012)	-0.046*** (0.013)
Chinese	-0.111*** (0.006)	-0.129*** (0.006)	Subjects allied to medicine	-0.006** (0.003)	-0.006** (0.003)
Other Asian	-0.110*** (0.004)	-0.124*** (0.005)	Biological sciences	-0.013*** (0.002)	-0.014*** (0.002)
Mixed	-0.043*** (0.002)	-0.047*** (0.003)	Veterinary science	-0.063* (0.038)	-0.067 (0.042)
Other ethnic group	-0.085*** (0.005)	-0.095*** (0.006)	Agriculture & related subjects	-0.075*** (0.006)	-0.081*** (0.007)
Socio-economic background (parental occupation)			Physical sciences	-0.076*** (0.003)	-0.082*** (0.003)
Higher managerial/professional	+	+	Mathematical sciences	-0.117*** (0.004)	-0.127*** (0.004)
Lower managerial/professional	-0.008*** (0.001)	-0.009*** (0.001)	Computer science	0.013*** (0.003)	0.014*** (0.003)
Intermediate	-0.009*** (0.002)	-0.010*** (0.002)	Engineering & technology	-0.020*** (0.003)	-0.021*** (0.003)
Small employers/own account workers	-0.012*** (0.002)	-0.013*** (0.002)	Architecture, building & planning	-0.017*** (0.003)	-0.018*** (0.004)
Technical/lower supervisory	-0.015*** (0.002)	-0.016*** (0.002)	Law	-0.031*** (0.003)	-0.033*** (0.003)
Semi-routine	-0.022*** (0.002)	-0.024*** (0.002)	Business & administrative studies	0.023*** (0.002)	0.024*** (0.002)
Routine	-0.033*** (0.002)	-0.036*** (0.002)	Mass communications & documentation	0.035*** (0.003)	0.036*** (0.003)
Long-term unemployed/Never worked	-0.045*** (0.012)	-0.049*** (0.013)	Languages	0.025*** (0.002)	0.026*** (0.003)
Unknown	-0.018*** (0.001)	-0.020*** (0.002)	Historical & philosophical studies	0.050*** (0.002)	0.051*** (0.002)
Parental education			Creative arts & design	0.013*** (0.002)	0.013*** (0.002)
Parents without HE qualifications	+	+	Education	-0.030*** (0.003)	-0.032*** (0.003)
Parents with HE qualifications	0.006*** (0.001)	0.006*** (0.001)	Combined subject	-0.004** (0.002)	-0.004** (0.002)
Unknown/Refused	-0.012*** (0.001)	-0.013*** (0.001)	***Mode of study***		
Student's characteristics			Part-time	+	+
Male	-0.053*** (0.001)	-0.057*** (0.001)	Full-time	0.313*** (0.005)	0.350*** (0.005)

Continued on next page

Table 3.5. (continued)

Age	-0.012*** (0.001)	-0.013*** (0.001)	Sandwich	0.379*** (0.005)	0.416*** (0.005)
Disability	-0.022*** (0.001)	-0.023*** (0.001)	Other	0.006 (0.007)	0.007 (0.008)
Home fees eligible	0.036*** (0.006)	0.038*** (0.007)	***Region of university***		
Term-time accommodation			London	+	+
Own residence	+	+	North East	-0.063*** (0.002)	-0.065*** (0.002)
Parental/guardian home	-0.003* (0.002)	-0.003* (0.002)	North West	-0.054*** (0.002)	-0.055*** (0.002)
Provider's property	-0.010*** (0.002)	-0.010*** (0.002)	Yorkshire	-0.050*** (0.002)	-0.051*** (0.002)
Private-sector halls	-0.013*** (0.003)	-0.014*** (0.003)	East Midlands	-0.040*** (0.002)	-0.041*** (0.002)
Other rented	0.009*** (0.002)	0.009*** (0.002)	West Midlands	-0.022*** (0.002)	-0.022*** (0.002)
Other	-0.004 (0.003)	-0.005 (0.003)	East of England	-0.024*** (0.002)	-0.025*** (0.002)
Not in attendance	0.032*** (0.011)	0.033*** (0.011)	South East	-0.023*** (0.002)	-0.024*** (0.002)
Unknown	-0.002 (0.002)	-0.002 (0.002)	South West	-0.034*** (0.002)	-0.034*** (0.002)
Major source of tuition fees			Wales	-0.077*** (0.002)	-0.080*** (0.002)
No award/backing	+	+	Scotland	-0.173*** (0.003)	-0.189*** (0.003)
UK LEA award	-0.004*** (0.001)	-0.004*** (0.001)	N. Ireland	-0.091*** (0.003)	-0.096*** (0.004)
Provider waiver	0.004 (0.010)	0.005 (0.010)	***Length of programme***		
UK central government	-0.023*** (0.004)	-0.024*** (0.004)	<= 2 years	+	+
Other	-0.006 (0.005)	-0.006 (0.005)	2-3 years	-0.059*** (0.005)	-0.062*** (0.006)
Pre-entry characteristics			3-4 years	0.029*** (0.006)	0.029*** (0.006)
Tariff Score	0.001*** (0.000)	0.001*** (0.000)	4-20 years	0.071*** (0.006)	0.069*** (0.006)
Public school	+	+	***Academic year***		
Private school	-0.009*** (0.001)	-0.010*** (0.002)	2010/11	+	+
Unknown school type	0.030*** (0.003)	0.031*** (0.003)	2011/12	0.019*** (0.001)	0.021*** (0.002)
Distance travelled (km)	0.0002*** (0.000)	0.0002*** (0.000)	2012/13	0.032*** (0.001)	0.034*** (0.001)
			2013/14	0.046*** (0.001)	0.048*** (0.001)
			2014/15	0.050*** (0.001)	0.053*** (0.001)
Observations	970,858				
Pseudo R^2	0.112				

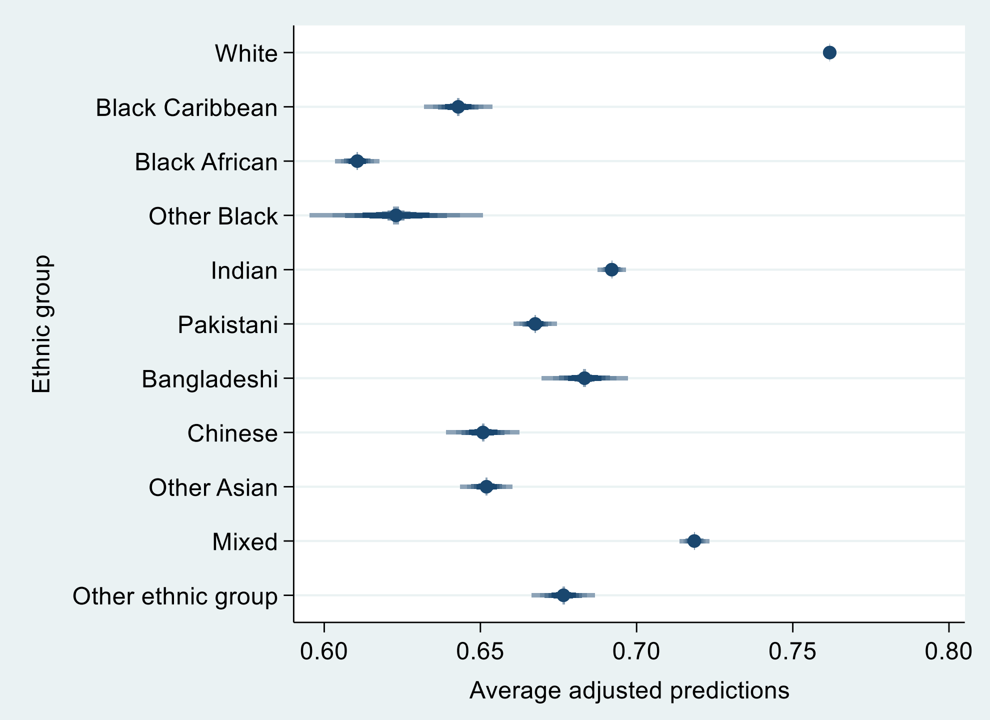
Note: The marginal effects shown in the table are derived from the original logistic regression (post-estimates).
*+ Reference category. Standard errors in parentheses based on the delta method. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*
*The model includes the following independent variables and interaction terms: "ethnicity", "age", "male", "disability", "home fees eligible", "socio-economic classification", "parental education", "subject of study", "mode of study", "length of programme", "type of university", "region of university", "major source of tuition fees", "term-time accommodation", "tariff score", "type of school", "distance travelled", "academic year", "ethnicity*male", "ethnicity*type of university", "ethnicity*socio-economic classification", and "ethnicity*tariff score". The cases with unknown ethnicity (<1% of the initial sample) are dropped from the regression analysis. For the dummy variables with a significant proportion of missing (unknown) values (>5%), I have included an additional category ("unknown").*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.6. Average adjusted predictions (AAPs) and adjusted predictions at means (APMs)

	Dependent variable: Good degree	
Ethnic group	**AAPs**	**APMs**
White	0.762***	0.793***
	(0.000)	(0.001)
Black Caribbean	0.643***	0.665***
	(0.006)	(0.007)
Black African	0.611***	0.625***
	(0.004)	(0.004)
Other Black	0.623***	0.638***
	(0.014)	(0.017)
Indian	0.692***	0.715***
	(0.002)	(0.003)
Pakistani	0.668***	0.687***
	(0.004)	(0.004)
Bangladeshi	0.683***	0.703***
	(0.007)	(0.008)
Chinese	0.651***	0.664***
	(0.006)	(0.006)
Other Asian	0.652***	0.669***
	(0.004)	(0.005)
Mixed	0.718***	0.746***
	(0.002)	(0.003)
Other ethnic group	0.677***	0.698***
	(0.005)	(0.006)
Observations	970,858	
Pseudo R^2	0.112	

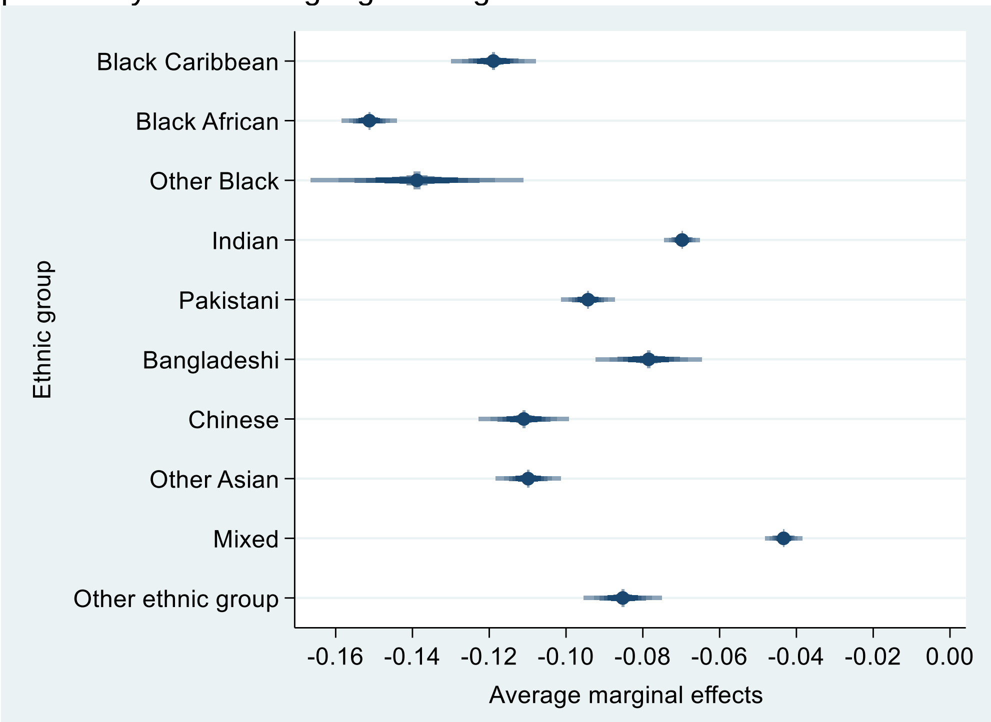
Note: The predicted probabilities shown in the table are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5.
Standard errors in parentheses based on the delta method.
** p < 0.1, ** p < 0.05, *** p < 0.01*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 3.2. Average adjusted predictions (probability of obtaining a good degree) by ethnic group



Note: The predicted probabilities shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. The shaded areas represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 3.3. Average marginal effects of ethnic groups on the probability of obtaining a good degree



Note: The average marginal effects (AMEs) shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. For each ethnic minority group, the figures represent the difference in the conditional likelihood of getting a good degree relative to White students (reference category). The shaded areas represent the 95% confidence intervals of the estimated marginal effects.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

3.4.2 Interaction effects

Using both tables and graphic visualisations, this subsection explores the interaction effects between ethnicity and students' gender, social class, type of university attended, and previous qualifications on the probability of graduating with a good degree.

Starting with the gender side of intersectionality, the results show that there are no significant gender differences in the magnitude of ethnic gaps among Black students (Table 3.7). This suggests that although all Black minorities are less likely to achieve good grades at university than their White peers (Figure 3.4), these gaps do not depend on their gender. On the contrary, there are statistically significant differences in the extent of ethnic disparities between men and women across all Asian communities, except for Chinese. Specifically, within the group of students who identify as Asian, the ethnic gaps in higher education attainment are more prominent amongst women than men. The difference is particularly notable for Bangladeshi students, whose underperformance (relative to White students) is 4 percentage points lower for men than women. Nevertheless, students from "other ethnic group" backgrounds witness the most significant disparities between genders (4.7 percentage points favouring men).

There is also a remarkable degree of interdependence between ethnicity and type of higher education institution in the likelihood of attaining a good degree (Table 3.8 and Figure 3.5). Black students perform better in the Russell Group universities than in the post-1992 HE providers, as the attainment gap compared to their White counterparts is significantly smaller in these 24 selective institutions. For example, the ethnic difference in the probability of graduating with first-class or upper second-class honours for an average Black Caribbean student enrolled at a Russell Group university (relative to his/her White peers) is 5.6 percentage points lower than that of a Black Caribbean counterpart attending a post-1992 university. The literature suggests that students from ethnic minorities are under-represented in "old" universities (Coffield and Vignoles, 1997; Boliver, 2013, 2016). Under the mounting political and social pressure for providing equal opportunities and support, the Russell Group and other "old" institutions have likely taken initiatives and targeted actions towards tackling racial inequalities and improving the attainment and overall higher education experience of Black minorities.

However, the story for students from Asian backgrounds is contrasting. Within these minorities, ethnic gaps are more prominent in the Russell Group universities than the other institution types. The picture is more disappointing for Bangladeshi students, as their ethnic gap in academic attainment in Russell Group institutions is about twice as large as in the post-1992 HE providers (10.2 versus 5.7 percentage points). These results suggest that high-status universities need to expand their policy interventions to Asian minorities, whereas post-1992 institutions should focus more on Black students to understand and reduce ethnic disparities in academic performance.

Table 3.9 shows how the effect of ethnicity on the dependent variable changes across different levels of students' socio-economic background (see also Figure 3.6). For both Black African and Black Caribbean minorities, there are no statistically significant differences in the relative size of their underachievement, irrespective of whether they come from "high-class", "middle-class" or "working-class" families. However, the ethnic gaps in attainment for students who describe themselves as "other Black" deteriorate as we move from higher to lower segments of the social class distribution. Specifically, the ethnic gap for "other Black" graduates doubles from 10.3 percentage points for those whose parents are employed in higher managerial/professional occupations to 21.4 percentage points for those in routine professions.

Interestingly, the impact of socio-economic background on the predicted probabilities of gaining a good degree works in the opposite direction for all Asian students. More specifically, the attainment gap of these students (compared to their White peers) is reduced markedly in the bottom levels of the socio-economic distribution relative to the top social class segment. For example, the ethnic gap in attainment for Indian and Pakistani minorities is 5.2 percentage points smaller for students from low socio-economic backgrounds (routine occupations) than those from a high social class (higher managerial/professional occupations). The corresponding difference is even larger amongst Chinese students, standing at nearly 7 percentage points. This is likely associated with cultural capital, work ethics and the value attached to education, as Chinese families from lower social classes might see higher education as a route to success and development to a greater extent than other ethnic minorities of similar socio-economic backgrounds (Francis and Archer, 2005). In contrast, there are no statistically significant differences in the size of ethnic gaps for the Mixed and "Other" ethnic groups (albeit these groups are complex and heterogeneous).

Finally, Figure 3.7 graphically shows how the probability of graduating with a good degree varies across selected values[16] of previous achievement (*tariff score*). On average, White students are more likely to achieve good grades across all levels of the prior educational ability distribution, and the differences are more apparent (and statistically significant) when the tariff score is greater than 100. However, the curve slope is not constant for all ethnic groups suggesting heterogeneous effects of prior attainment on the likelihood of gaining a good degree.

Table 3.7. Interaction effects of gender and ethnicity on the probability of obtaining a good degree

Ethnic group	Ethnic gap in Pr(Good degree)		Second difference
	Women	Men	
Black Caribbean	-0.116***	-0.123***	0.007
Black African	-0.154***	-0.148***	-0.006
Other Black	-0.134***	-0.145***	0.012
Indian	-0.075***	-0.063***	-0.012***
Pakistani	-0.104***	-0.082***	-0.022***
Bangladeshi	-0.096***	-0.056***	-0.040***
Chinese	-0.111***	-0.111***	0.000
Other Asian	-0.122***	-0.094***	-0.028***
Mixed	-0.041***	-0.046***	0.005
Other ethnic group	-0.106***	-0.059***	-0.047***

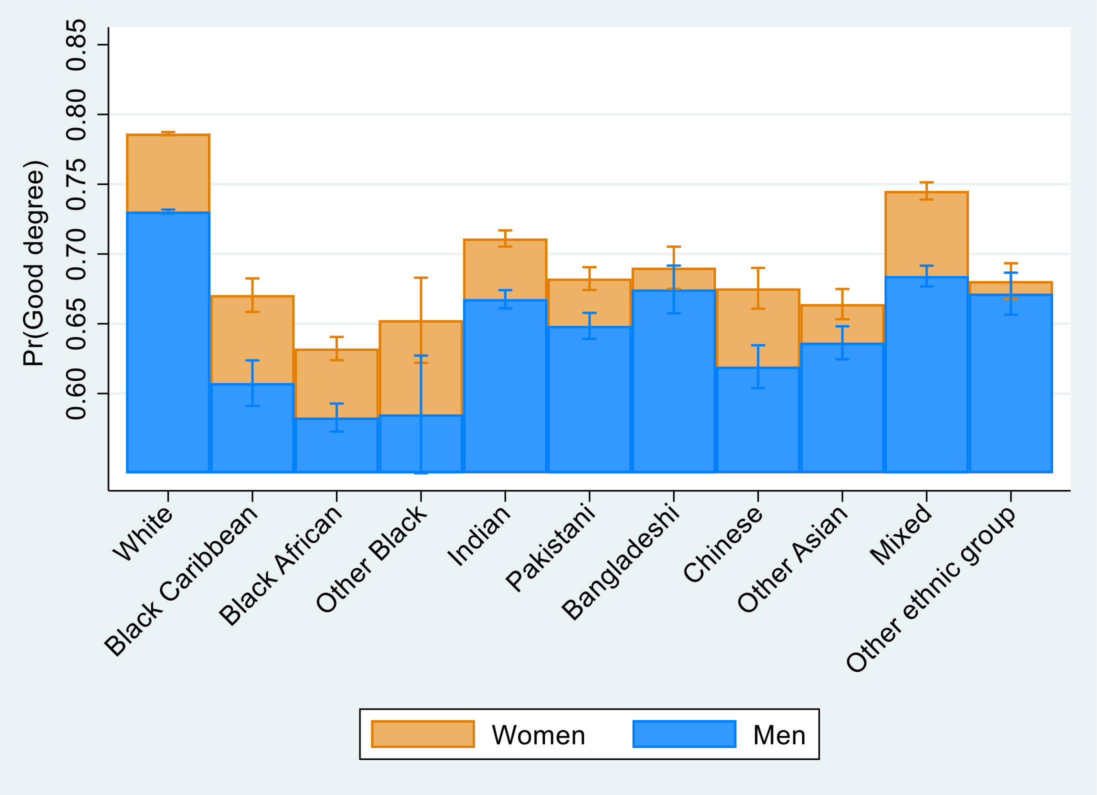
Note: The coefficients within each gender (ethnic gap) represent the difference in the average probability of obtaining a good degree between each ethnic minority and White students (reference category). The rightmost column shows the difference in the ethnic gap between men and women (i.e., second difference). The results are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5.

* *p < 0.1,* ** *p < 0.05,* *** *p < 0.01*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

[16] *Because* tariff score *is a continuous variable, in Figure 3.7, I present specific values of its distribution to explore the interactive effect of* "ethnicity*tariff score*". These values correspond to representative points of interest with many observations. For 97% of graduates, the* tariff score *value is less than 600 points, whereas approximately one-third scored below 300 points. The mean value of the* tariff score *in the regression sample is 353.*

Figure 3.4. Probability of obtaining a good degree by gender and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. The error bars represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.8. Interaction effects of university type and ethnicity on the probability of obtaining a good degree

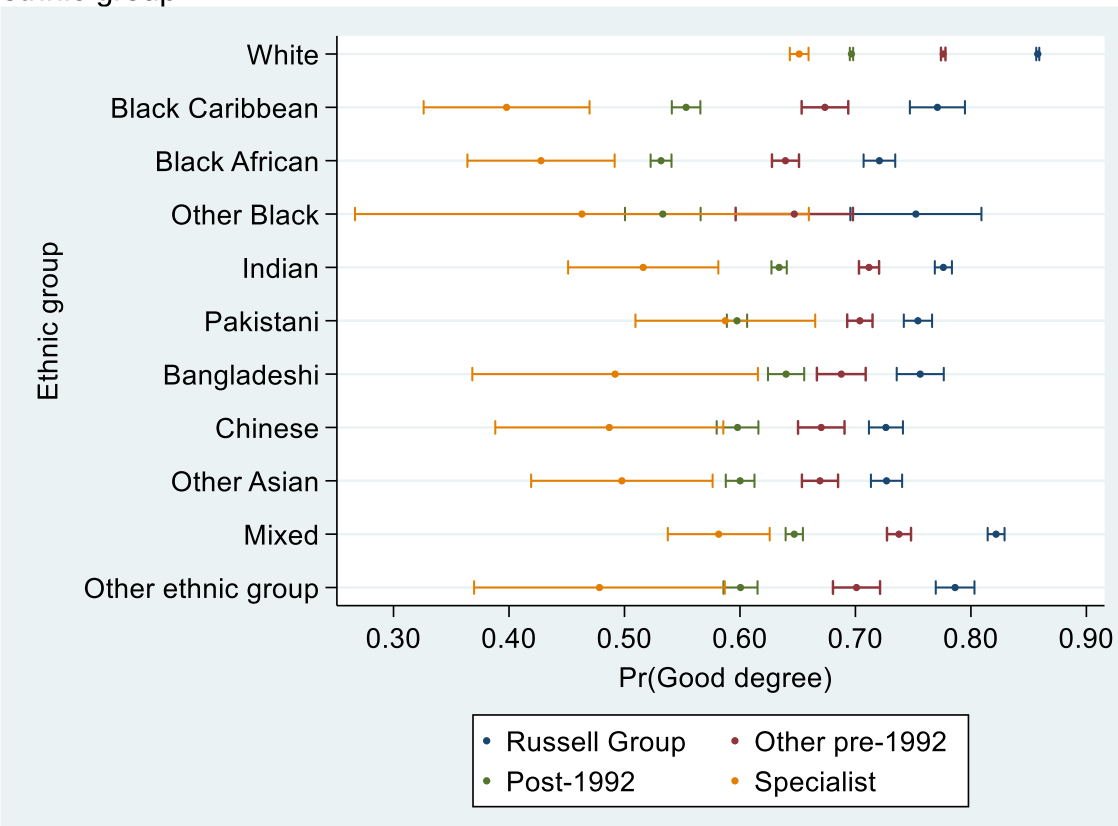
Ethnic group	Ethnic gap in Pr(Good degree)			Second difference		
	Russell Group (1)	Other pre-1992 (2)	Post-1992 (3)	(1-2)	(1-3)	(2-3)
Black Caribbean	-0.087***	-0.102***	-0.143***	0.016	0.056***	0.041***
Black African	-0.137***	-0.137***	-0.165***	-0.001	0.028***	0.028***
Other Black	-0.106***	-0.129***	-0.163***	0.023	0.058*	0.034
Indian	-0.082***	-0.064***	-0.063***	-0.017***	-0.019***	-0.002
Pakistani	-0.104***	-0.072***	-0.099***	-0.032***	-0.005	0.027***
Bangladeshi	-0.102***	-0.088***	-0.057***	-0.013	-0.045***	-0.032***
Chinese	-0.131***	-0.106***	-0.099***	-0.026**	-0.033***	-0.007
Other Asian	-0.131***	-0.107***	-0.096***	-0.024**	-0.035***	-0.010
Mixed	-0.036***	-0.038***	-0.050***	0.002	0.013**	0.011
Other ethnic group	-0.072***	-0.075***	-0.096***	0.004	0.025**	0.021

Note: The coefficients within each university type (ethnic gap) represent the difference in the average probability of obtaining a good degree between each ethnic minority and White students (reference category). The "second difference" columns show the difference in the ethnic gap between types of institutions. The results are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. Specialist institutions are not presented in the table as they cover a small proportion of students (1.4%).

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 3.5. Probability of obtaining a good degree by university type and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. The error bars represent the 95% confidence intervals of the estimated predicted probabilities.

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.9. Interaction effects of socio-economic background and ethnicity on the probability of obtaining a good degree

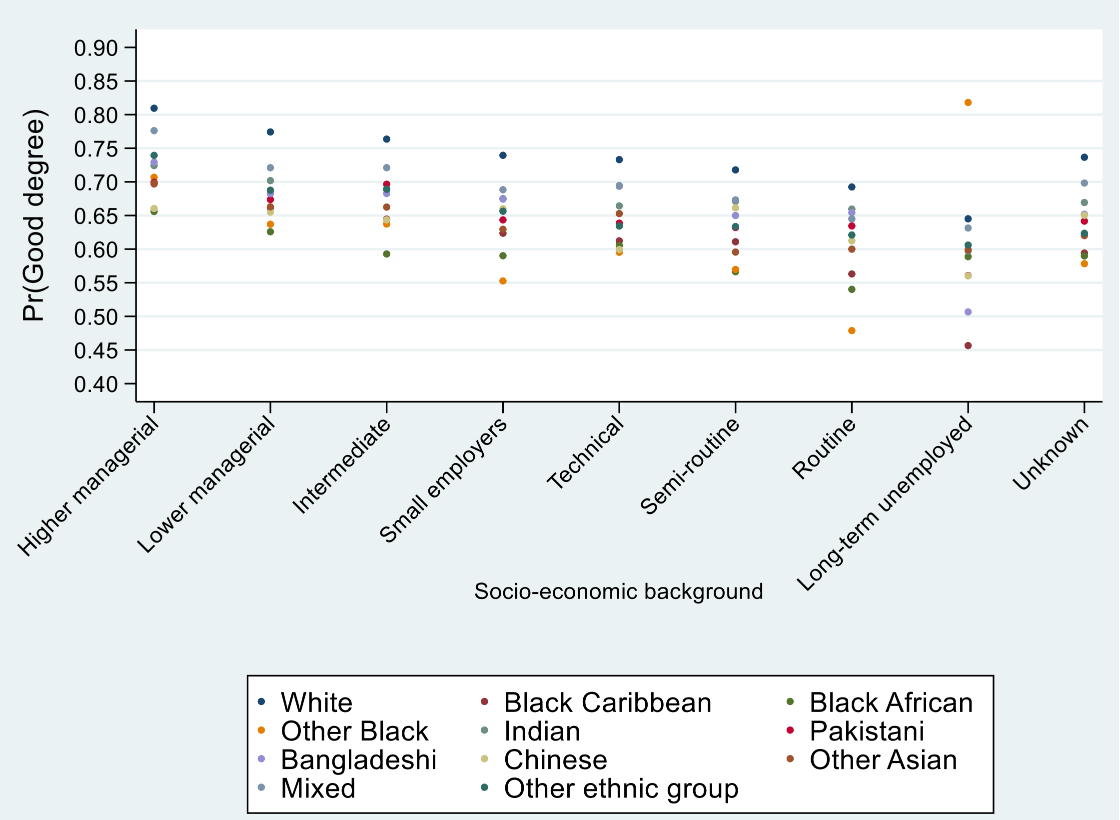
Ethnic group	**Ethnic gap in Pr(Good degree)**			**Second difference**		
	Higher managerial (1)	**Small employers (2)**	**Routine (3)**	**(1-2)**	**(1-3)**	**(2-3)**
Black Caribbean	-0.113***	-0.116***	-0.129***	0.003	0.017	0.014
Black African	-0.154***	-0.149***	-0.152***	-0.004	-0.001	0.003
Other Black	-0.103***	-0.187***	-0.214***	0.084	0.111**	0.027
Indian	-0.085***	-0.064***	-0.033***	-0.021**	-0.052***	-0.032***
Pakistani	-0.110***	-0.096***	-0.058***	-0.014	-0.052***	-0.038***
Bangladeshi	-0.080***	-0.065***	-0.038**	-0.015	-0.042	-0.027
Chinese	-0.149***	-0.080***	-0.080***	-0.070***	-0.069**	0.001
Other Asian	-0.112***	-0.110***	-0.092***	-0.002	-0.019	-0.018
Mixed	-0.033***	-0.051***	-0.047***	0.018	0.014	-0.004
Other ethnic group	-0.070***	-0.083***	-0.071***	0.013	0.001	-0.012

Note: The coefficients within each level of socio-economic background (ethnic gap) represent the difference in the average probability of obtaining a good degree between each ethnic minority and White students (reference category). The "second difference" columns show the difference in the ethnic gap between socio-economic levels. The results are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5.

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

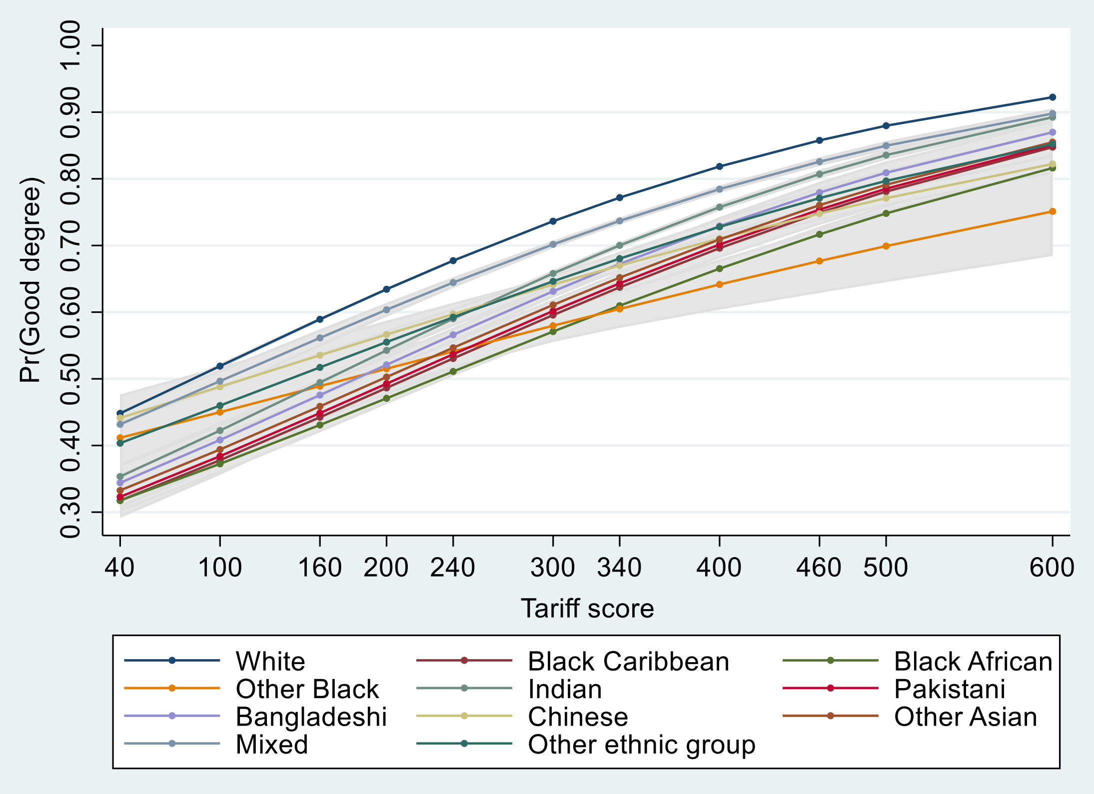
Figure 3.6. Probability of obtaining a good degree by socio-economic background and ethnic group



Note: The predicted probabilities shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. For clarity, the confidence intervals of the predicted probabilities are not plotted in the graph, as they overlap across ethnic groups.

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Figure 3.7. Probability of obtaining a good degree by ethnic group at selected levels of tariff score



Note: The predicted probabilities shown in the graph are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. The shaded areas represent the 95% confidence intervals of the estimated predicted probabilities.
Source: HESA (pooled data for the academic years 2010/11–2014/15)

3.4.3 Robustness checks

3.4.3.1 Peer effects, university quality measures, and additional interaction terms

I performed a series of robustness checks on the results presented above. First, I included some additional peer effects and university quality variables and interaction terms in the logistic regression models. Peer effects may partially determine undergraduates' performance via their social and academic interaction, while institutional factors (such as teaching effectiveness and university quality and selectivity) could also influence the academic outcomes. Specifically, the new variables are:

"Relative tariff score": This variable is the fraction of a student's tariff score over the average tariff score of his/her peers. It measures the students' relative academic ability. I computed students' average tariff score within the same university, course, and academic year. I also incorporated the quadratic term of the "*relative tariff score*" variable to capture the potential differential impact of

academic peer effects on the probability of graduating with a good degree at various levels of prior educational performance (Johnes and McNabb, 2004).

"Proportion of non-White peers": I calculated this variable within the same university, course, and academic year to control for the likely effect of ethnic diversity on students' attainment. This variable is related to students' integration into their department's environment and partly captures the impact of social mix on academic performance.

"University's average tariff score": This variable is a measure of university selectivity and shows the average institution's tariff score (at entry) for each academic year. Selective institutions might affect students' academic experience differently from other universities with respect to structures and student support services.

"Staff-student ratio": Within each university and academic year, this variable represents the ratio of academic staff (including atypical) to the total number of students. I collected these figures from the publicly available HESA student and staff records and merged them with the main HESA datasets. This variable is a measure of teaching quality.

"Non-White/White staff ratio": Within each university and academic year, this variable represents the ratio of non-White academic staff (including atypical) to White academic staff. This ratio partially captures the ethnic minorities' "sense of belongingness" to the university and the likely effect of role models on their attainment (UUK and NUS, 2019).

"University's income per student": This is the fraction of each institution's total income over the number of its students. It is a measure of university quality. I collected these figures from the HESA finance records and merged them with the primary HESA datasets.

"Teaching Excellence Framework (TEF) outcome": This variable shows the award (Gold, Silver or Bronze) given to the UK higher education providers. It is a measure of universities' quality for their undergraduate provision. I obtained the TEF outcomes from the Office for Students, as updated in June 2020 (OfS, 2020). The universities' engagement in the TEF assessment is currently voluntary.

*"Ethnicity*subject of study"*: The effect of students' ethnicity on academic performance may vary across courses. I incorporated this additional interaction

term in the models to capture the potential differential effect of ethnicity on the probability of gaining a good degree across subjects of study.

*"Ethnicity*academic year"*: This interaction term allows ethnic gaps in academic attainment to vary across different academic years.

*"Type of university*academic year"*: This product term controls for the fact that some types of institutions may exhibit higher grade inflation than others during the period considered (Bachan, 2015).

As Table 3.A3 in the Appendix shows, controlling for the additional factors outlined above makes little difference to the size of ethnic gaps in academic performance relative to the initial results (Table 3.5). In fact, the differences in the likelihood of obtaining a good degree (relative to the White ethnic group) slightly increase by approximately 1-2 percentage points for most ethnic minorities (except for students from Chinese and "other Black" backgrounds). This implies that the effect of being from an ethnic minority group on the probability of achieving a good degree had a small upward bias (toward zero) before adjusting for those extra variables and product terms.

3.4.3.2 University and course fixed effects

The university quality measures introduced above might not encapsulate other institution-related characteristics (such as university structures, the level of academic support, and pastoral care) that could influence the academic performance differently across ethnic groups. Therefore, the second robustness check incorporates university and subject of study fixed effects in the models. This approach estimates the effect of ethnicity on the dependent variable conditional on individuals studying in the same subject area and institution. Therefore, this within-university technique allows the time-invariant unobserved institution- and course-specific components to be correlated with the explanatory variables. Because of the high dimension of the fixed effects (given the large number of universities and courses), I switch to the linear probability model (LPM) to overcome convergence issues in the logit fixed-effects models (Guimaraes and Portugal, 2010). Moreover, I cluster the standard errors by university, considering that the dependent variable's unobserved determinants should be correlated for students at the same institution.

Table 3.A4 in the Appendix reveals that the results remain robust even to inclusion of "*university*course*" fixed effects. For most ethnic minority groups, the magnitude of ethnic gaps does not decline compared to the main logistic

regression specification. Instead, accounting for university and subject of study fixed effects marginally increases the ethnic differences (by 0.3 to 1.2 percentage points) in the propensity of graduating with a good degree between ethnic minorities and White students. These minimal differences should partly reflect the different techniques used (that is, the logistic regression versus the LPM). Notwithstanding, the unobserved university- and course-related determinants are not a key driver of the ethnic gaps documented in the present study.

3.4.3.3 Exploring the other side of interaction effects

The third robustness test refers to how the interaction effects are calculated in practice when using average marginal effects in logistic regression models. The results reported in subsection 3.4.2 (Tables 3.7-3.9) correspond to the interaction effects of independent variables (gender, university type, and socio-economic background) at different categories of ethnicity (which is treated as fixed). Nevertheless, in non-linear models, one should also investigate the other side of interdependence by testing the interaction effect of ethnicity at various levels of the other regressors (Mize, 2019). Table 3.A5 in the Appendix shows that there are some differences in the size of ethnic gaps between the two approaches, especially concerning the interaction effect of university type and ethnicity. However, these differences are small in most cases, while the direction and statistical significance of the interaction effects remain unchanged.

3.5 Conclusion

This paper uses personal-level data from the HESA, focusing on all UK-domiciled first-degree students who graduated from UK universities during the academic years 2010/11–2014/15. The results confirm earlier research findings that all ethnic minorities have substantially lower chances of achieving first-class or upper-second class honours (that is, a “good degree”) than White students (Broecke and Nicholls, 2007; Richardson, 2008). The present paper substantially improves the existing literature by using recent data, enriched with comprehensive information on prior attainment, parents’ social class and other characteristics. Unlike most studies to date, this work also estimates interaction effects between ethnicity and students’ gender, social class, university type, and previous attainment on the probability of being awarded a good degree.

The present analysis allows the probability of achieving a good degree to vary according to a wide range of determinants of higher education performance: socio-demographic traits (such as gender, age, disability status, socio-economic background, and parental education); institutional and study characteristics (type and region of university, subject of study, mode of study, length of programme); pre-entry factors (prior educational ability, type of school, distance travelled); and other individual-level characteristics (term-time accommodation, home fees eligibility and source of tuition fees).

Despite including all these variables, the ethnic gaps in achievement favouring White students remain large and statistically significant. On average, the probability of attaining a good degree stands at 76.2% for White students, and the difference relative to ethnic minorities ranges from 4.3 percentage points (for students from a Mixed ethnic background) to 15.1 percentage points (for Black African students). The picture is particularly disappointing for the Black community and remains worrying for all Asian minorities (Chinese, Pakistani, Bangladeshi, Indian, and other Asian students). The findings in this paper suggest that under-attainment by ethnic minorities is a multifaceted and pervasive issue. Ethnic gaps are remarkable across all types of institutions (Russell Group universities, other pre-1992, and post-1992 institutions) and social class levels. In addition, the choice of the subject of study and previous educational attainment do not eliminate the existing ethnic disparities.

This study provides a detailed mapping of ethnic discrepancies in university attainment by analysing the interdependency between ethnicity and specific characteristics. There are notable dissimilarities in the extent of ethnic gaps between men and women across all Asian communities, except for Chinese. Specifically, the inequalities in higher education attainment for these ethnic groups (especially for Bangladeshi students) are more prominent amongst women than men. On the other hand, the magnitude of ethnic gaps for Black students does not depend on their gender. On average, the attainment gap for Black students (compared to their White peers) is significantly smaller in the Russell Group institutions than the rest types of higher education providers (other pre-1992 and post-1992 universities). In contrast, the underperformance of Asian students is more pronounced in the Russell Group universities. For example, the average difference in the probability of graduating with a good degree between White and Bangladeshi students is about double the size in Russell Group institutions compared to the post-1992 HE providers (10.2 versus 5.7 percentage points).

Moreover, the attainment gap for Asian students decreases as we move from the top to the bottom level of social class. For instance, the difference in attainment for Chinese minorities (compared to their White fellows) is 7 percentage points smaller for students from low socio-economic backgrounds (routine professions) than those from a high socio-economic background (higher managerial/professional occupations). On the contrary, there are no statistically significant disparities in the extent of ethnic gaps for the Black African, Black Caribbean, Mixed, and “Other” ethnic minorities according to their social class. The above findings may serve as a compass for policymakers to design targeted measures that will remove racial inequalities in attainment rather than implement general interventions applicable to all ethnic minorities and universities.

Another interesting finding of the present work is the consistency in the rank ordering of ethnic gaps in the likelihood of gaining a good degree across various institution types, social classes, and genders. Specifically, conditional on all other observable characteristics, the ethnic effect on academic performance is ranked uniformly (for most ethnic groups) within each type of university, gender, and socio-economic group (see Figures 3.4, 3.5 and Table 3.9). Similarly, the ordering of ethnic groups (relative to White students) is generally preserved over the previous attainment distribution, although the ethnic gaps in attainment vary as we move to higher levels of prior ability. This rank ordering of White students ahead of all ethnic minorities likely suggests that the ethnic gaps are predominantly driven by some structural mechanisms that translate ethnicity into poorer outcomes, rather than unobserved background characteristics that are common across all ethnic minorities and in contrast to the White group. Such structural barriers that compromise ethnic minorities’ academic performance may be linked to institutional culture and support systems, rigid course content and delivery, and racial discrimination in formal assessments.

Achieving high grades at the undergraduate level is a typical prerequisite for pursuing postgraduate studies (Masters or Doctorate degrees). A good degree also unlocks opportunities in the labour market and improves earnings and career prospects (Naylor, Smith and Telhaj, 2016; Feng and Graetz, 2017). It is, therefore, critical to ameliorate the inequitable outcomes in academic performance amongst ethnic groups. The recent Government’s initiatives, which, among other measures, oblige all universities to publish attainment figures separately by ethnic group (DfE, 2019b), are a step in the right direction.

However, tackling racial inequalities in higher education may require further collaboration between national policymakers and universities to build a holistic approach and instil a universal cultural change.

Policies aiming to remove ethnic inequalities in higher education attainment should focus on developing inclusive learning environments, which would provide equal support, guidance, and opportunities to all students. According to Stevenson's (2012) study, many interviewed students from ethnic minority backgrounds reported a lack of academic support in their studies. Moreover, based on the outcomes of a "student retention and success programme", which ran across 13 universities and various disciplines in the UK (Thomas et al., 2017), ethnic minorities have a lower feeling of "belongingness" to their institution than their White British peers. Similarly, an older survey exploring students' higher education experiences highlighted that these groups express lower levels of satisfaction and a sense of anxiety and isolation at their university (NUS, 2011).

In this context, a recent report commissioned by the Universities UK and the NUS describes the best practices and actions taken by higher education institutions to improve the overall university experience of ethnic minorities (UUK and NUS, 2019). It also provides specific recommendations to universities about creating an ethnically diverse environment, which would enhance the awareness of cultural differences. In addition, Singh (2011) acknowledges the "preparedness for success" as a significant determinant of academic achievement. This term includes prior experiences in primary and secondary education, material prosperity, cultural capital, and parental support, which all impact students' success.

Ridley (2007) inferred that some ethnic minorities use less efficient methods of studying, which subsequently affect their marks in the exams. For example, she found that Black students demonstrate more frequent "surface approaches" in their learning style than their White peers. The so-called "deficit model" suggests that any differences in attainment are attributable to the students' characteristics (such as skills and experience), and it neglects the impact of institutional structures or discrimination. However, in the present analysis, I show that the dramatic differences in academic performance between ethnic minorities and White students are not totally explained by their diverse socio-economic or educational/ability profile.

Although this study includes most of the known variables that affect higher education performance, it does not disentangle the effects of ethnicity from other unobserved factors that are still likely to influence the probability of achieving good grades at university. Some determinants of attainment described above (such as the different learning styles, self-motivation, academic support, and discrimination), which are likely to differ between ethnic groups, are not observed in administrative datasets. As a result, the extent of these variables' effects on the academic performance of ethnic minorities remains unknown.

Further research must better understand the hurdles and specific cultural differences that compromise ethnic minorities' achievement and higher education experience. In this context, performing qualitative and focus group interviews with a representative sample of university students and staff across different UK regions might shed more light on the unknown parameters of ethnic minorities' underperformance. Specifically, the relevant questionnaires could cover sections related to institutional structures, students' socio-demographic characteristics, attainment, learning styles, and decision-making associated with their choices when looking ahead at graduation. The adoption of a comprehensive mixed-method approach by exploiting qualitative and quantitative data and methods is likely to provide valuable insights that would assist policymakers in addressing the higher education underachievement of students from ethnic minority backgrounds in the UK.

References of Chapter 3

Agresti, A., 2013. *Categorical Data Analysis.* 3rd ed. New York: Wiley.

Ai, C. and Norton, E.C., 2003. Interaction terms in logit and probit models. *Economics Letters,* 80(1), pp. 123-129.

Allison, P.D., 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research,* 28(2), pp. 186-208.

Amemiya, T., 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature,* 19(4), p. 1483.

Bachan, R., 2015. Grade inflation in UK higher education. *Studies in Higher Education,* 42(8), pp. 1580-1600.

Berry, W.D., Demeritt, J.H.R. and Esarey, J., 2010. Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential? *American Journal of Political Science,* 54(1), pp. 248-266.

BIS, 2014. *National strategy for access and student success in higher education.* London: BIS (Department for Business Innovation and Skills).

Boliver, V., 2013. How fair is access to more prestigious UK universities? *British Journal of Sociology,* 64(2), pp. 344-364.

Boliver, V., 2016. Exploring Ethnic Inequalities in Admission to Russell Group Universities. *Sociology,* 50(2), pp. 247-267.

Bradley, S. and Taylor, J., 2004. Ethnicity, educational attainment and the transition from school. *Manchester School,* 72(3), pp. 317-346.

Breen, R., Karlson, K.B. and Holm, A., 2018. Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. *Annual Review of Sociology, Vol 44,* 44(1), pp. 39-54.

Broecke, S. and Nicholls, T., 2007. *Ethnicity and degree attainment.* London: Department for Education and Skills, (Research Report RW92).

Buis, M.L., 2010. Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal,* 10(2), pp. 305-308.

Cameron, C. and Trivedi, P., 2005. *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

Christie, H., 2007. Higher Education and Spatial (Im)Mobility: Nontraditional Students and Living at Home. *Environment and Planning A,* 39(10), pp. 2445-2463.

Coffield, F. and Vignoles, A., 1997. *Widening participation in higher education by ethnic minorities, women and alternative Students, Report 5, Higher Education in the Learning Society.* London: NCIHE (National Committee of Inquiry into Higher Education).

Connor, H., La Valle, I., Tackey, N.D. and Perryman, S., 1996. *Ethnic minority graduates: Differences by degrees.* Brighton: The Institute for Employment Studies (IFE), (309).

Connor, H., Tyers, C., Modood, T. and Hillage, J., 2004. *Why the difference? A closer look at higher education minority ethnic students and graduates.* Nottingham: DfES (Department for Education and Skills), (RR552).

Cotton, D.R.E., Joyner, M., George, R. and Cotton, P.A., 2016. Understanding the Gender and Ethnicity Attainment Gap in UK Higher Education. *Innovations in Education and Teaching International,* 53(5), pp. 475-486.

Crawford, C., 2014. *The link between secondary school characteristics and university participation and outcomes.* London: Department for Education (Centre for Analysis of Youth Transitions), (DFE-RR353).

Crawford, C. and Greaves, E., 2015. *Socio-economic, ethnic and gender differences in HE participation.* London: BIS (Department for Business Innovation and Skills), (186).

DfE, 2019a. *Participation Rates In Higher Education: Academic Years 2006/2007 – 2017/2018 (Provisional).* London: DfE (Department for Education).

DfE, 2019b. Universities must do more to tackle ethnic disparity. *Department for Education, The Cabinet Office* [Online]. Available from: https://www.gov.uk/government/news/universities-must-do-more-to-tackle-ethnic-disparity [Accessed 20 December 2020].

Dowd, B.E., Greene, W.H. and Norton, E.C., 2014. Computation of standard errors. *Health services research,* 49(2), p. 731.

Feng, A. and Graetz, G., 2017. A question of degree: The effects of degree class on labor market outcomes. *Economics of Education Review,* 61, pp. 140-161.

Fielding, A., Charlton, C., Kounali, D. and Leckie, G., 2008. *Degree attainment, ethnicity and gender: Interactions and the modification of effects A quantitative analysis.* York: Higher Education Academy.

Francis, B. and Archer, L., 2005. British–Chinese pupils' and parents' constructions of the value of education. *British Educational Research Journal,* 31(1), pp. 89-108.

Greaves, E., Macmillan, L. and Sibieta, L., 2014. *Lessons from London schools for attainment gaps and social mobility.* London: Social Mobility and Child Poverty Commission.

Guimaraes, P. and Portugal, P., 2010. A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal,* 10(4), pp. 628-649.

HEFCE, 2010. *Student Ethnicity: Profile and progression of entrants to full-time, first degree study.* Bristol: HEFCE (Higher Education Funding Council for England), (2010/13).

HEFCE, 2018. *Differences in student outcomes-The effect of student characteristics.* Bristol: HEFCE (Higher Education Funding Council for England), (2018/05).

Johnes, G. and McNabb, R., 2004. Never Give up on the Good Times: Student Attrition in the UK. *Oxford Bulletin of Economics and Statistics,* 66(1), pp. 23-47.

Khambhaita, P. and Bhopal, K., 2015. Home or away? The significance of ethnicity, class and attainment in the housing choices of female university students. *Race Ethnicity and Education,* 18(4), pp. 535-566.

Leslie, D., 2005. Why people from the UK's minority ethnic communities achieve weaker degree results than whites. *Applied Economics,* 37(6), pp. 619-633.

Long, J.S. and Freese, J., 2014. *Regression models for categorical dependent variables using Stata.* 3rd ed. Texas: Stata Press.

Long, J.S. and Mustillo, S.A., 2018. Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes. *Sociological Methods & Research*, pp. 1-37.

Meunier, M., de Coulon, A., Marcenaro-Gutierrez, O. and Vignoles, A., 2013. A longitudinal analysis of UK second-generation disadvantaged immigrants. *Education Economics,* 21(2), pp. 105-134.

Mize, T.D., 2019. Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science,* 6, pp. 81-117.

Modood, T., 2003. Ethnic differentials in educational performance. In: D. Mason, ed. *Explaining Ethnic Differences: Changing Patterns of Disadvantage in Britain.* Bristol: ESRC & The Policy Press.

Mood, C., 2009. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review,* 26(1), pp. 67-82.

Naylor, R., Smith, J. and Telhaj, S., 2016. Graduate returns, degree class premia and higher education expansion in the UK. *Oxford Economic Papers,* 68(2), pp. 525-545.

Naylor, R.A. and Smith, J., 2004. Determinants of Educational Success in Higher Education. In: G. Johnes and J. Johnes, eds. *International Handbook on the Economics of Education.* Cheltenham, UK: Edward Elgar Publishing.

NUS, 2011. *Race for Equality: A report on the experiences of Black students in further and higher education.* London: NUS (National Union of Students).

OfS, 2020. *TEF outcomes* [Online]. OfS (Office for Students),. Available from: https://www.officeforstudents.org.uk/advice-and-guidance/teaching/tef-outcomes/#/tefoutcomes/ [Accessed 15 June 2020].

Owen, D., Green, A., Pitcher, J. and Maguire, M., 2000. *Minority ethnic participation and achievements in education, training and the labour market.* London: Department for Education and Employment, (225).

Platt, L., 2007. *Poverty and Ethnicity in the UK.* Bristol: The Policy Press.

Rainey, C., 2016. Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction. *Political Science Research and Methods,* 4(3), pp. 621-639.

Richardson, J.T.E., 2008. The attainment of ethnic minority students in UK higher education. *Studies in Higher Education,* 33(1), pp. 33-48.

Richardson, J.T.E., 2015. The Under-Attainment of Ethnic Minority Students in UK Higher Education: What We Know and What We Don't Know. *Journal of Further and Higher Education,* 39(2), pp. 278-291.

Richardson, J.T.E., Mittelmeier, J. and Rienties, B., 2020. The role of gender, social class and ethnicity in participation and academic attainment in UK higher education: an update. *Oxford Review of Education,* 46(3), pp. 346-362.

Ridley, A., 2007. Approaches to learning, age, ethnicity and assessment. Implications for widening participation. *Psychology Teaching Review,* 13(1).

Shiner, M. and Modood, T., 2002. Help or Hindrance? Higher Education and the Route to Ethnic Equality. *British Journal of Sociology of Education,* 23(2), pp. 209-232.

Singh, G., 2011. *Black and minority ethnic (BME) students' participation in higher education: improving retention and success.* York: Higher Education Academy.

Stevenson, J., 2012. *Black and minority ethnic student degree retention and attainment.* York: The Higher Education Academy.

Strand, S., 2015. *Ethnicity, Deprivation and Educational Achievement at Age 16 in England: Trends Over Time: Annex to Compendium of Evidence on Ethnic Minority Resilience to the Effects of Deprivation on Attainment.* London: DfE (Department for Education), (DFE-RR439B).

Thomas, L., Hill, M., O'Mahony, J. and Yorke, M., 2017. *Supporting student success: strategies for institutional change. What Works? Student Retention & Success programme.* London: HEFCE - Higher Education Academy - Paul Hamlyn Foundation - Action on Access. Available from: *https://www.phf.org.uk/publications/works-student-retention-success-full/* [Accessed 20 December 2020].

UUK and NUS, 2019. *Black, Asian and minority ethnic student attainment at UK universities: #Closingthegap.* London: Universities UK (UUK) and National Union of Students (NUS). Available from: *https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2019/bame-student-attainment-uk-universities-closing-the-gap.pdf* [Accessed 20 December 2020].

Vidal Rodeiro, C. and Zanini, N., 2015. The role of the A grade at A level as a predictor of university performance in the United Kingdom. *Oxford Review of Education,* 41(5), pp. 647-670.

Williams, R., 2012. Using the margins command to estimate and interpret adjusted predictions and marginaleffects. *Stata Journal,* 12(2), pp. 308-331.

Wilson, D., Burgess, S. and Briggs, A., 2011. The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics,* 24(2), pp. 681-700.

Appendix of Chapter 3

Table 3.A1. Description of the variables used in this paper

Variable	Values	Description – Notes
Age	17-53 years	Age of student on the 31^{st} of August in the reported academic year.
Class of first degree	First-class honours, Upper second-class honours, Third-class honours/Pass	This variable is included in Table 3.1 only. Some qualifications related to medical studies (such as medicine & dentistry and veterinary science) have a huge proportion of unclassified degrees. Therefore, these subjects of study represent a relatively small number of observations in this analysis sample.
Disability	Yes, No	A binary variable showing whether a student has reported a disability.
Distance travelled	0-987 kilometers	This variable shows the distance (in a straight line) between student's home postcode (as reported before entry) and the institution's main campus location.
Ethnic group	White, Black Caribbean, Black African, Other Black, Indian, Pakistani, Bangladeshi, Chinese, Other Asian, Mixed, Other ethnic group	Student's self-reported ethnicity. I adopt the recommended classification based on the 2011 National Statistics. The White category includes graduates from White British, Irish, Gypsy or Irish Traveller, and any other White backgrounds. Although the White category contains some ethnic minorities, the inconsistent way institutions record its sub-categories across different countries or regions of the UK does not allow me to distinguish White British from the rest of White students. However, in line with the practice adopted in other studies in the field, the "ethnic minority" term used throughout this paper comprises only non-White graduates. "Other ethnic group" includes Arab minorities and any other group not mentioned in any of the rest categories. HESA does not provide clarifications about the "Other Black" category. This group likely includes students who identify as "Black European" or "Black North American".
Good degree	Yes, No	A binary variable capturing students who attained either a first-class or an upper second-class honours degree.
Home fees eligible	Yes, No	A binary variable showing whether a student is entitled to pay "home" tuition fees.
Length of programme	<= 2 years, 2-3 years, 3-4 years, 4-20 years.	Represents the expected duration (in years) of the programme, from the start of study to the course's end. I dropped from the analysis very few observations with an unknown duration or an expected programme length over 20 years.
Major source of tuition fees	No award/financial backing, UK LEA mandatory award, Provider waiver, UK central government and Local Authorities, Other	Represents the student's primary source of tuition fees. "UK LEA mandatory award" comprises cases where the "Student Loans Company" (or the "Student Awards Agency" for Scotland) covers either the entire amount of tuition fees or a part of them (and students paying the remaining share). "Other" category includes charities, research councils, UK industries or student's employer, international agencies, and other overseas foundations.

Continued on next page

Table 3.A1. (continued)		
Male	Yes, No	A binary variable capturing male graduates.
Mode of study	Part-time, Full-time, Sandwich, Other mode of study	"Part-time" includes graduates who studied on courses with a duration of fewer than 24 weeks per academic year and evening students. "Sandwich" covers students who attended a thin or thick sandwich course with study or placement amounting to at least 21 hours/week for no less than 24 weeks/academic year.
Parental education	Parents without HE qualifications, Parents with HE qualifications, Unknown/refused	This captures graduates who reported that at least one of their parents/guardians holds a higher education qualification.
Region of university	London, North East, North West, Yorkshire, East Midlands, West Midlands, East of England, South East, South West, Wales, Scotland, N. Ireland	The location of higher education institution based on the postcode of its main administration premises.
Socio-economic classification	Higher managerial & professional occupations, Lower managerial & professional occupations, Intermediate occupations, Small employers & own account workers, Lower supervisory & technical occupations, Semi-routine occupations, Routine occupations, Never worked & long-term unemployed, Unknown/Not classified	This represents the occupation of student's parent/guardian with the highest earnings.
Subject of study	Medicine & dentistry, Subjects allied to medicine, Biological sciences, Veterinary science, Agriculture & related subjects, Physical sciences, Mathematical sciences, Computer science, Engineering & technology, Architecture, building & planning, Social studies, Law, Business & administrative studies, Mass communications & documentation, Languages, Historical & philosophical studies, Creative arts & design, Education, Combined subject	The subject area of the first-degree course based on the 19 principal codes of the "Joint Academic Coding System" (JACS) classification. "Combined" subjects include joint degrees in one or over one code of subject, irrespective of the percentage contribution of each subject area (e.g., History (60%) & Politics (40%), Economics (90%) & Mathematics (10%)).
Subject of study (grouped areas)	STEM, LEM, Other subject, Combined subject	The STEM category covers subjects related to Science, Technology, Engineering, and Mathematics, as well as Architects and Health subjects. The LEM category includes subjects in Law, Economics, and Management. The remaining subjects (except for the combined ones) are grouped in the "Other" category. This grouping is only presented in Table 3.A2 of the Appendix.

Continued on next page

Table 3.A1. (continued)

Tariff score	5-1,991 points	This is an aggregated score from student's prior qualifications. During the application process, the "Universities and Colleges Admissions Service" (UCAS) computes each student's total tariff points based on his/her qualifications and then provides them to HESA. This variable approximates the student's prior educational ability. The tariff score of 99% of the graduates included in this study is less than 680 points, whereas only 0.9% of them have less than 80 points.
Term-time accommodation	Own residence, Parental/guardian home, Provider's property, Private-sector halls, Other rented, Other, Not in attendance, Unknown	This variable defines the place where a student lived during the reported academic year. "Other rented" category refers to temporary arrangements (e.g., yearly house share). "Not in attendance" category captures students who were not in attendance at the university during the reported academic year because of industrial placement or other reasons (e.g., language year abroad).
Type of school	Public school, Private school, Unknown school type	Indicates the type of the previous provider attended by a student before entering higher education. "Public school" includes state-funded schools and colleges.
Type of university	Russell Group universities, Other pre-1992 institutions, Post-1992 institutions, Specialist institutions	Russell Group universities include 24 prestigious institutions. The 1992 Further and HE Act and following legislation resulted in abolishing the so-called "binary divide" between the centrally funded universities and the locally funded polytechnics. As a consequence, over 40 former polytechnics were granted degree-award status after 1992. The "Post-1992" category includes these former polytechnics along with the institutions established after 1992. "Specialist" institutions cover a small fraction of the analysis sample and mainly refer to colleges/universities of agriculture, arts, music and drama, etc.

Source: HESA

Table 3.A2. Mean characteristics: White versus non-White graduates

Variable	White	non-White	Difference
Good degree	0.749	0.612	0.14^{***}
Subject of study area			
STEM	0.339	0.361	-0.02^{***}
LEM	0.183	0.282	-0.10^{***}
Other subject	0.289	0.152	0.14^{***}
Combined	0.190	0.205	-0.02^{***}
Type of university			
Russell Group	0.291	0.223	0.07^{***}
Other Pre-1992	0.206	0.216	-0.01^{***}
Post-1992	0.484	0.549	-0.07^{***}
Specialist	0.019	0.011	0.01^{***}
Region of HE provider			
London	0.077	0.327	-0.25^{***}
North East	0.060	0.018	0.04^{***}
North West	0.129	0.092	0.04^{***}
Yorkshire	0.118	0.077	0.04^{***}
East Midlands	0.089	0.105	-0.02^{***}
West Midlands	0.071	0.124	-0.05^{***}
East of England	0.048	0.073	-0.03^{***}
South East	0.131	0.106	0.02^{***}
South West	0.098	0.036	0.06^{***}
Wales	0.068	0.019	0.05^{***}
Scotland	0.081	0.022	0.06^{***}
N. Ireland	0.031	0.002	0.03^{***}
Mode of Study			
Part-time	0.016	0.038	-0.02^{***}
Full-time	0.901	0.885	0.02^{***}
Sandwich	0.074	0.059	0.01^{***}
Other mode of study	0.009	0.018	-0.01^{***}
Student's characteristics			
Male	0.440	0.435	0.00^{***}
Age	20.762	20.910	-0.15^{***}
Disability	0.115	0.076	0.04^{***}
Home fees eligible	0.992	0.982	0.01^{***}
Term-time accommodation			
Own residence	0.096	0.097	-0.00
Parental/guardian home	0.213	0.420	-0.21^{***}
Provider's property	0.080	0.078	0.00^{**}
Other/unknown accommodation	0.611	0.405	0.21^{***}
Socio-economic background (parental occupation)			
Higher managerial/professional	0.226	0.139	0.09^{***}
Lower managerial/professional	0.265	0.207	0.06^{***}
Intermediate	0.111	0.090	0.02^{***}
Small employers/own account workers	0.060	0.083	-0.02^{***}
Technical	0.045	0.024	0.02^{***}
Semi-routine	0.085	0.142	-0.06^{***}
Routine	0.041	0.057	-0.02^{***}
Long-term unemployed/Never worked	0.001	0.004	-0.00^{***}
Unknown occupation	0.165	0.255	-0.09^{***}
Parental education			
Parents with HE qualifications	0.448	0.370	0.08^{***}
Parents without HE qualifications	0.322	0.368	-0.05^{***}
Unknown parental education	0.230	0.262	-0.03^{***}
Pre-entry characteristics			
Tariff Score	359.345	322.366	36.98^{***}
Private school	0.118	0.088	0.03^{***}
Public school	0.845	0.862	-0.02^{***}
Unknown school type	0.037	0.050	-0.01^{***}
Distance travelled (in km)	112.196	72.347	39.85^{***}
Observations	**933,880**	**212,655**	**-**

Note: The equality of means between the two groups is examined using standard tests of proportions. The numbers of observations are rounded to the nearest multiple of 5, in line with data provider's disclosure control.

**p<0.1, **p<0.05, ***p<0.01.*

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.A3. Robustness analysis – Logistic regression: Average marginal effects (AMEs) and Marginal effects at means (MEMs)

	Dependent variable: Good degree	
Ethnic group	**AMEs**	**MEMs**
White	+	+
Black Caribbean	-0.125*** (0.006)	-0.135*** (0.007)
Black African	-0.164*** (0.004)	-0.183*** (0.005)
Other Black	-0.135*** (0.016)	-0.149*** (0.019)
Indian	-0.083*** (0.003)	-0.092*** (0.003)
Pakistani	-0.115*** (0.004)	-0.128*** (0.005)
Bangladeshi	-0.101*** (0.008)	-0.113*** (0.009)
Chinese	-0.107*** (0.007)	-0.124*** (0.007)
Other Asian	-0.120*** (0.005)	-0.135*** (0.006)
Mixed	-0.047*** (0.003)	-0.050*** (0.003)
Other ethnic group	-0.093*** (0.006)	-0.103*** (0.007)
Observations	960,914	
Pseudo R^2	0.113	

Note: The marginal effects shown in the table are derived from the logistic regression (post-estimates). Standard errors in parentheses based on the delta method. In the logistic regression, robust standard errors are used (that is, the variance estimator is robust to certain misspecification forms).
+Reference category.
* p < 0.1, ** p < 0.05, *** p < 0.01
*The model specification of this robustness analysis contains all the variables and interaction terms of the logistic regression in Table 3.5, the additional interaction terms “ethnicity*subject of study”, “ethnicity*academic year”, “type of university*academic year”, and the additional variables capturing the peer effects and university quality (that is, “relative tariff score” and its squared term, “proportion of non-White peers”, “university’s average tariff score”, “academic staff-student ratio”, “non-White/White staff ratio”, “university’s income per student”, “Teaching Excellence Framework (TEF) outcome”). Unlike the sample of the regression presented in Table 3.5 of the main text, students who graduated from medicine & dentistry, veterinary, and agricultural subjects are excluded from the present models. The reason is that the number of students within these subjects is relatively small (for example, most medicine and veterinary degrees are unclassified). Hence, when I attempted to include those subjects in the current specification, the logistic model did not converge in the sample, and the coefficients of ethnicity could not be estimated. The cases with unknown ethnicity (<1% of the initial sample) are dropped from the regression analysis. For the dummy variables with a significant proportion of missing (unknown) values (>5%), I have included an additional category (“unknown”).*
Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.A4. Robustness analysis – Linear probability model (LPM) with subject of study and university fixed effects: Marginal effects

	Dependent variable: Good degree
Ethnic group	**AMEs**
White	+
Black Caribbean	-0.128*** (0.007)
Black African	-0.159*** (0.006)
Other Black	-0.151*** (0.018)
Indian	-0.080*** (0.004)
Pakistani	-0.104*** (0.005)
Bangladeshi	-0.088*** (0.009)
Chinese	-0.110*** (0.008)
Other Asian	-0.116*** (0.006)
Mixed	-0.046*** (0.003)
Other ethnic group	-0.090*** (0.005)
Observations	970,839
Adjusted R^2	0.143

Note: The marginal effects shown in the table are derived from the linear probability model (post-estimates). Standard errors in the LPM are clustered by university, considering that the unobserved factors for students belonging to the same institution should be correlated. The average marginal effects (AMEs) and the marginal effects at the means (MEMs) are the same for linear regressions.

+Reference category.

* p < 0.1, ** p < 0.05, *** p < 0.01

*The model specification of this robustness analysis contains "subject of study*university" fixed effects. It also includes all the variables and interaction terms of the logistic regression in Table 3.5, except for the "subject of study", "type of university", and "region of university", because the latter variables are collinear with the fixed effects. For the same reason, the present model specification excludes most of the additional peer effects and university quality variables incorporated in Table 3.A3 of the robustness analysis section (that is, "proportion of non-White peers", "university's average tariff score", "academic staff-student ratio", "non-White/White staff ratio", "university's income per student", "Teaching Excellence Framework (TEF) outcome"). However, it includes the "relative tariff score" and its squared term.*

The cases with unknown ethnicity (<1% of the initial sample) are dropped from the regression analysis. For the dummy variables with a significant proportion of missing (unknown) values (>5%), I have included an additional category ("unknown").

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Table 3.A5. Robustness analysis - Interaction effects on the probability of obtaining a good degree (second difference approach)

Ethnic group	Ethnicity*gender	Ethnicity*university type			Ethnicity*socioeconomic background		
	Women – Men	Russell group – Other pre-1992	Russell group – Post-1992	Other pre-1992 – Post-1992	Higher managerial – Small employers	Higher managerial – Routine	Small employers – Routine
Black Caribbean	0.016*	0.001	0.035*	0.034***	-0.014	-0.003	0.011
Black African	-0.001	-0.016	0.008	0.024***	-0.018	-0.017	0.002
Other Black	0.025	0.031	0.081**	0.050	0.079	0.103*	0.024
Indian	-0.013***	-0.036***	-0.046***	-0.010	-0.027***	-0.055***	-0.028**
Pakistani	-0.019***	-0.050***	-0.026***	0.024***	-0.023*	-0.061***	-0.037***
Bangladeshi	-0.040***	-0.033**	-0.073***	-0.040***	-0.016	-0.038	-0.022
Chinese	-0.002	-0.022*	-0.016	0.006	-0.065***	-0.063**	0.002
Other Asian	-0.028***	-0.046***	-0.065***	-0.018*	-0.005	-0.019	-0.014
Mixed	0.006	-0.003	0.007	0.009	0.013	0.008	-0.005
Other ethnic group	-0.043***	-0.001	0.019	0.020	0.005	-0.008	-0.013

Note: The columns show the difference in the ethnic gaps in the probability of obtaining a good degree between genders, type of institutions and socio-economic background (i.e., "second differences"). The results are derived from the original logistic regression (post-estimates), based on the same variables and interaction terms as in Table 3.5. Specialist institutions are not presented in the table as they cover a small proportion of students (1.4%). The ethnic gap is defined as the difference in the average likelihood of obtaining a good degree between each ethnic minority and White students (reference category).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: HESA (pooled data for the academic years 2010/11–2014/15)

Statement of Authorship

<table>
<tr><td colspan="10">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="10">Labour market inequalities amongst UK-born university graduates: What drives wage differentials between ethnic groups?</td></tr>
<tr><td colspan="10">Publication status (tick one)</td></tr>
<tr><td>Draft manuscript</td><td>✓</td><td>Submitted</td><td></td><td>In review</td><td></td><td>Accepted</td><td></td><td>Published</td><td></td></tr>
<tr><td>Publication details (reference)</td><td colspan="9"></td></tr>
<tr><td colspan="10">Copyright status (tick the appropriate statement)</td></tr>
<tr><td colspan="3">I hold the copyright for this material</td><td>✓</td><td colspan="5">Copyright is retained by the publisher, but I have been given permission to replicate the material here</td><td></td></tr>
<tr><td>Candidate's contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="9">The candidate contributed to / considerably contributed to / predominantly executed the…

Formulation of ideas: 95%

Design of methodology: 100%

Experimental work: N/A

Presentation of data in journal format:100%</td></tr>
<tr><td>Statement from Candidate</td><td colspan="9">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td colspan="5">Konstantinos Kollydas</td><td colspan="2">Date</td><td colspan="2">24/04/2021</td></tr>
</table>

4. Labour market inequalities amongst UK-born university graduates: What drives wage differentials between ethnic groups?

Abstract

In light of the reignited debate in the media about ethnic pay inequalities, this study examines whether UK-born university graduates from Black, Asian, and Mixed/Other ethnic backgrounds experience wage penalties in the UK labour market. This paper draws on restricted data from the Annual Population Survey (2013-2018), which has been recently enriched with information about respondents' higher education (the subject of first degree and type of institution attended). The model results provide firm evidence that wage inequalities persist even after allowing for differences in a comprehensive set of higher education characteristics, demographic traits, and occupation-related factors. Wage gaps are strikingly more pronounced for Black employees, standing at 16.7% for men and 4.5% for women compared to equally qualified White workers, and hold up against a series of robustness checks and sensitivity methods, suggesting racial discrimination. The study also finds heterogeneous effects of ethnicity on wages across different segments of the earnings distribution and within subgroups of workers. The decomposition technique outcomes reveal that characteristics associated with the employees' occupation and industry sector (such as the under-representation of ethnic minorities in high-paid jobs) account for half of the earnings differential between White and non-White individuals.

Keywords: wage gaps, ethnic minorities, discrimination, decomposition analysis, unobservable selection

JEL classification: I24, J31, J71

4.1 Introduction

4.1.1 Background and research objectives

There has been a heated debate in the UK media about earnings disparities between people from ethnic minority backgrounds and their White counterparts (see, for example, Croxford's and Topham's (2018) articles in the *BBC* and *The Guardian*, respectively). Despite the principles against racial discrimination in employment, as set out by the UK legislation (e.g., Race Relations Act, 1976; Equality Act, 2010) and other policy initiatives, ethnic inequalities seem to persist in British society. In October 2018, the former Prime Minister Theresa May considered launching a scheme that would mandate businesses to publish figures regarding "ethnicity wage gaps" (Swinford, 2018). The reporting of pay gaps among ethnic groups is not compulsory yet in the UK business world, although it is expected to become law within the next few years. Nonetheless, some large firms and organisations (including the Bank of England, Citigroup, Deloitte, Ernst & Young, KPMG, and others) have published wage differences amongst ethnic groups in recent years. For instance, Deloitte UK (2019) reported an average hourly pay gap of 12.9% for its Black, Asian and minority ethnic (BAME) employees relative to non-BAME individuals.

However, exclusively looking at raw figures may paint an incomplete picture of the earnings differentials and the underlying mechanisms driving ethnic pay gaps because of concealed dissimilarities in employees' characteristics. Most importantly, little is known about whether ethnic pay inequalities exist among higher education (HE) graduates (Zwysen and Longhi, 2018). This is essential to examine because university graduates, who are highly qualified and accumulate large human capital levels, represent a rapidly growing share of the UK workforce. To fill this gap, this paper explores the extent of wage differences between UK-born university graduates from Black, Asian, or Mixed/Other minority backgrounds and their White counterparts. As I discuss below, by confining the sample to UK-born graduates, I attempt to provide a more precise interpretation of the unexplained part of earnings gaps, thus directing policymaking to targeted interventions.

Using the UK Annual Population Survey (APS) data for years 2013-2018, firstly, I provide a comprehensive description of ethnic pay inequalities between

UK-born, first-degree holders by presenting the unconditional ("raw") hourly wage gaps among the four ethnic groups and descriptively investigating the main wage determinants. Secondly, I estimate the conditional ("adjusted") ethnic differences in wages for men and women separately, after controlling for an extensive set of human capital characteristics, occupational and demographic traits that affect the level of earnings. Thirdly, I examine heterogeneous effects of ethnicity on wages, according to the type of university attended, subject area of study, degree class, major occupation group, workplace size, and across age bands. On the same wavelength, I use quantile regression methods to inspect whether ethnic pay gaps vary across different wage distribution levels. Fourthly, by employing a decomposition technique, I quantify the contribution of each main group of observable characteristics to the earnings differential between White and non-White workers. Fifthly, I examine the robustness of the OLS results by adopting a partial identification approach (Oster, 2017), which estimates the degree of selection on unobservable characteristics required to cancel the effects of ethnicity identified in OLS regressions. I also employ the most common matching methods used in the literature to address issues concerning the OLS functional form assumptions and the selection on observed characteristics.

The 1991 Census brought in a question regarding the ethnic background (on a self-ascribed basis) of people living in the UK, which was subsequently incorporated into many government surveys. Heath and others (e.g., Cheng and Heath, 1993) introduced the concept of "ethnic penalties" in recognition of inequalities in occupational attainment between ethnic minority groups and British people with similar qualifications and background characteristics. Several studies in sociology and economics systematically explore ethnic penalties in the UK labour market and pinpoint discrepancies in wages or the probability of employment experienced by certain ethnic minorities (e.g., Blackaby et al., 1998; Berthoud, 2000; Lindley, Dale and Dex, 2004; Heath and Cheung, 2006; Longhi and Platt, 2008; Algan et al., 2010; Li and Heath, 2010; Rafferty, 2012; Modood and Khattab, 2016). A smaller amount of research extends this to the ethnoreligious grounds by investigating ethnic disadvantages along the lines of religion (see, for example, Lindley, 2002a; Khattab and Johnston, 2013; Longhi, Nicoletti and Platt, 2013; Khattab and Johnston, 2015). Other works concentrate on the labour market experience of first generation immigrants (Dustmann and

Fabbri, 2005; Clark and Drinkwater, 2008), but they do not explicitly consider the effect of ethnicity (Price, 2001).

This work advances prior research in two ways. First, by using the most recent APS data analysed within the restricted Secure Lab of the UK Data Service, I exploit the information related to the subject of the first degree and university attended, which became available to researchers only after 2012. Unlike studies to date, by incorporating the major of bachelor's degree and the type of institution into the analysis, I allow for heterogeneity in the distribution of these variables across ethnic groups and account for their impact on wages. In doing so, I acknowledge the recent evidence from the UK, which signifies diverse labour market returns to higher education according to the degree subject and university attended (Britton et al., 2016; Walker and Zhu, 2018; Belfield et al., 2018a, 2018b).

Second, this research sheds light on earnings differentials amongst UK-born graduates who received their first degree from a UK university. To the best of my knowledge, this is the first time in the UK context that a study explores ethnic differences in wages over the life cycle by exclusively focusing on this subpopulation of the workforce. The rationale behind restricting the sample to UK-born graduates is twofold. The rapid expansion in higher education has enormously improved the acquisition of human capital in the labour force. In the UK, the overall HE participation rate has risen dramatically since 1950. More specifically, the rate increased from 3.4% in 1950 to 12.4% in 1980 and 32% in 1995 (NCIHE, 1997). Subsequently, it followed an upward trend from 42% in 2006/07 to 50% in 2017/18 (DfE, 2019). Hence, analysing wage outcomes of employees with similar human capital (first-degree holders), who constitute a continuously growing proportion of the workforce, should be of particular interest for policymakers to better evaluate the factors that drive earnings inequalities and tackle the barriers that specific ethnic minority groups confront in the labour market. It is noteworthy that a government-commissioned report estimates an annual benefit of £24 billion (over 1% of the GDP) to the UK economy if Black and minority ethnic (BME) people become equally represented in the British labour market (McGregor-Smith, 2017).

The second reason for choosing the more homogeneous subgroup of employees mentioned above is that it helps control for some characteristics potentially important for labour market outcomes, which are not observed in the

APS data. Specifically, research reveals that language skills are a significant determinant of the labour market productivity and earnings (Lindley, 2002b; Dustmann and Fabbri, 2003; Miranda and Zhu, 2013; Yao and van Ours, 2015). Therefore, not controlling for English proficiency would introduce a downward bias in the estimated effect of ethnicity on wages if I included people born outside the UK in the analysis (especially those who migrated to Britain from a non-English speaking country after, say, the age of 12)[17]. Similarly, first-generation immigrants also encounter more severe obstacles than UK-born ethnic minorities regarding the economic and social assimilation, social networks, and the knowledge of the British labour market. These characteristics, which are rarely observed in administrative datasets, affect wages (Bandiera, Barankay and Rasul, 2009; Dustmann et al., 2016; Frattini, 2017).

Selection into HE and the ability bias pose further estimation concerns. For instance, it is well-established that differences between ethnic groups in HE participation rates are partially connected with cultural attitudes towards education (Dale et al., 2002; See et al., 2011). However, the APS datasets do not contain information about such characteristics (e.g., linguistic skills, cultural differences and individual aspirations towards HE, ability). Although differentials in ability almost certainly exist even amongst graduates, I partly alleviate such endogeneity issues by accounting for degree class, type of university, and subject area of the first degree. Employers often translate the degree class as a signal of ability, particularly at the early stages of employees' working life, where information on productivity before recruitment is limited (Naylor, Smith and Telhaj, 2016).

The traditional regression-based approaches study wage differences between ethnic groups by controlling for specific factors (such as education, experience, job tenure, and health status) that act as a proxy for individual productivity. If the coefficients on ethnicity variables remain statistically significant even after allowing for differences in the observed productivity-related determinants of wage, economists infer that there is evidence of discrimination (Darity Jr and Mason, 1998). However, in empirical projects, it is impossible to consider all potential productivity drivers that are valued in the job market and are

[17] The literature finds that age at immigration plays a significant role in immigrants' educational attainment and economic assimilation in the host country (e.g., Schaafsma and Sweetman, 2001; Clark and Lindley, 2008; Lemos, 2013).

likely to differ amongst ethnic groups. The unobserved factors that influence earnings include but are not limited to labour market attachment, social or professional network effects, economic motivation and career aspirations, negotiation skills, and pre-labour market characteristics (such as school quality and family socio-economic background).

In the present study, because of data limitations, it is difficult to disentangle the magnitude of racial discrimination from the above-mentioned unobserved variables that may also shape the level of earnings (Elliott and Lindley, 2008; Topa, 2011; Dustmann et al., 2016). Establishing a causal effect of each ethnic group on wages would require performing an experiment by randomly assigning ethnicity to individuals and subsequently measuring their wages. Such an experiment is ethically and practically infeasible to implement. The economic literature usually relies on conducting field experiments (such as audit and correspondence studies) to measure the extent of labour market discrimination precisely. For example, Bertrand and Mullainathan (2004) conducted a seminal field experiment in two cities of the United States (Chicago and Boston) by randomly distributing curricula vitae (CVs) to White- and Black-sounding names. They found that CVs with African-American names were 50% less likely to receive a favourable response (that is, a call-back or offer of work) by employers than White names.

However, unlike many non-experimental works that address wage inequalities, field experiment settings almost solely centre on the hiring process. As a result, there is very little evidence in the literature establishing how discrimination in the recruitment stage translates into wage gaps if applicants are hired (Pager, Western and Bonikowski, 2009). In this paper, by employing the partial identification technique proposed by Oster (2017), I calculate how large the effect of the unobservable (confounding) characteristics should be to eliminate the impact of ethnicity on wages (that is, to produce a zero coefficient of each ethnic minority group in the regression estimates after accounting for all observed wage determinants). Where the required effect of the unobservable factors is too extreme to be realistic, I conclude that there is clear evidence of racial discrimination against specific ethnic minority groups. Furthermore, I estimate the contribution of the observed covariates to the wage gaps between White and non-White employees by adopting a decomposition technique suggested by Neumark (1988).

4.1.2 Key findings

The regression model results show a staggering wage gap of 16.7% for Black male employees compared to their White counterparts, after controlling for a wide range of variables concerning demographic traits, higher education characteristics and job/sector factors. The ethnic penalties (in terms of hourly wages) are considerably smaller for Asian men (-4.1%) and statistically insignificant (-1.3%) for male employees from Mixed/Other ethnic backgrounds. Black and Asian women see substantially lower ethnic pay differences than men, standing at -4.5% and -2.0%, respectively. The latter figures are, to some extent, explained by the well-documented gender income inequality and discrepancies in occupational preferences between women and men.

The wage inequalities in favour of White employees persist (although in different magnitude) even within broader groups of subjects of study, university type, degree class, occupation, workplace size, and age, suggesting that the underlying dynamics (including discrimination and possible unobserved characteristics) that drive ethnic disadvantages constitute a scarring effect on the British labour market. A key finding of this study is that ethnic penalties among UK-born graduates increase with age. Specifically, there are no statistically significant wage differences for employees aged 30 and under across all ethnic groups (except Black men, who are penalised by 10.1% relative to similarly situated young White workers). On the contrary, substantial ethnic pay gaps exist for older employees (aged 31-65), ranging from 3.5% for Asian women to 19.4% for Black men. The overall picture is particularly concerning for the Black community, as the average life-course ethnic penalties for Black males remain unequivocally robust even after applying Oster's (2017) method described above, implying existence of racial discrimination.

By decomposing the wage differential into explained (30% of total wage gap) and unexplained (70%) components, I show that imbalances in job characteristics (such as the under-representation of ethnic minorities in high-salaried occupations and their shorter firm tenure) are responsible for half of the total wage differences between White and non-White employees, on average for both genders. These findings may provide a valuable direction for policymakers in better understanding the drivers that result in ethnic penalties in the UK labour market. For example, they could focus on initiatives for equal representation of comparably qualified employees from different ethnic backgrounds in the most privileged occupations (managerial/professional jobs), where participation is smaller for ethnic minorities, and wage gaps are wider (especially for Black employees).

This study proceeds as follows. In section 4.2, I discuss the previous research on ethnic penalties/discrimination in the labour market. Section 4.3 provides descriptive evidence and presents the data and the empirical strategy chosen for this work. Section 4.4 reports the findings from the regression and decomposition analyses, followed by the results from the sensitivity analysis in section 4.5. The last section summarises the findings and concludes by discussing the policy implications of this study.

4.2 Previous literature

There is a voluminous literature describing discrepancies in employment probability, economic activity, employment patterns and earnings between immigrants and British natives and across different ethnic groups (e.g., Blackaby et al., 1998; Berthoud, 2000; Lindley, Dale and Dex, 2004; Dustmann and Fabbri, 2005; Li and Heath, 2010; Khattab and Johnston, 2013). It should be noted that pay differences merely represent one of several labour market outcomes that ethnicity influences. This section mostly focuses on empirical studies that address earnings gaps between ethnic groups in the UK.

A significant body of research indicates that ethnic minorities, notably people from Black African/Caribbean, Bangladeshi and Pakistani backgrounds, are penalised in the UK labour market in terms of wages (see, for example, Carmichael and Woods, 2000; Heath and Cheung, 2006; Longhi and Platt, 2008; Algan et al., 2010; Brynin and Güveli, 2012; Longhi and Brynin, 2017; Li and Heath, 2018). Most of the existing literature on ethnic penalties relies on exploiting a sample of the working-age population, regardless of the level of employees' education and country of birth. For example, Brynin and Güveli (2012) use the Labour Force Survey (LFS) data for 1993-2008 to show that Pakistani, Bangladeshi, and Black African people face earnings penalties ranging from 0.07 to 0.21 log points compared to White employees. Similarly, Heath and Cheung (2006) exploit the LFS data for years 2001-2004 to illustrate that the earnings gaps for Bangladeshi, Black African and Pakistani men stand at -0.32, -0.24 and -0.14 log points, respectively (relative to comparably qualified White people). They find that the extent of ethnic penalties is smaller for Indian (0.06) and Chinese (0.07) men. The authors also reveal that the ethnic disadvantages are lower among women (compared to men), arguing that this reflects the gender pay disparities faced by British females rather than reduced discrimination against ethnic minority women.

Although these studies typically include controls for the proportion of graduates or the highest qualification in their regression analyses, they do not account for the subject of study and university type. One exception is the work of Zwysen and Longhi (2018). Using data from the “Destination of Leavers from Higher Education” survey for years 2004/05-2011/12, the authors restrict their sample to young university graduates who are British nationals to examine whether there exist ethnic pay disparities in early career stages (six months and three years after completing higher education). They show that there are only minor earnings differences between ethnic minorities and White British young graduates after controlling for job characteristics, parental background, local area factors, university type, degree classification, and subject of study. Specifically, the yearly earnings of Black Caribbean, Pakistani, and Bangladeshi females were 2.1%-4.4% lower than White British graduates six months after finishing university. In contrast, Indian, Bangladeshi, and Chinese men’s respective earnings were 2.1%-4.8% higher than White British employees. However, as I discuss in subsection 4.4.2, only focusing on the early employment stages, when there is remarkable noise in pay discrepancies and little wage dispersion, underestimates the lifetime ethnic pay gaps.

Unsurprisingly, most studies document that wage gaps are narrower for ethnic minority employees born in the UK than first-generation immigrants (Algan et al., 2010; Longhi and Brynin, 2017). The slower social and economic assimilation of first-generation immigrants in developed countries is a determining factor that leads to their disadvantages in terms of earnings, occupation compatibility and productivity compared to native employees (Chiswick, 1978; Shields and Price, 1998; Friedberg, 2000; Frattini, 2017). In the same context, education acquired abroad may be less valued in the labour market relative to qualifications obtained domestically (Valbuena and Zhu, 2018). Therefore, the present study compares UK-born employees to minimise the effect of such unobserved characteristics, which are likely to vary across ethnic groups.

Moreover, research shows that people from ethnic minority backgrounds are over-represented in lower-class occupations (e.g., Li, 2018). The present work confirms the latter findings and additionally estimates the gender-specific wage gap faced by non-White employees within broad occupation groups (upper-salaried and routine jobs). This becomes more insightful when considering the evidence of mismatching of education levels and occupations entered by ethnic minorities, a situation known as “over-education” (Rafferty, 2012).

It is also important to mention that the extent of ethnic penalties in wages addressed by empirical studies rarely captures the “pre-labour market disadvantage” faced by ethnic minorities (Heath and Cheung, 2006). Apart from discrimination and other unobserved personal factors, the wage gap possibly echoes skill differences between ethnic groups rooted in parental characteristics, school environments, and neighbourhood conditions. For instance, some ethnic minority groups live in disadvantaged areas with low school quality. As a result, ethnic minorities enter the labour market with unequal skill levels. This leads to an overestimation of the extent of wage gaps, even when researchers account for the number of schooling years in their analyses (Neal and Johnson, 1996). Although I do not explicitly control for these premarket factors in the regression specifications due to data limitations, I partially mitigate such endogeneity issues by constructing a homogeneous sample of equally educated employees (that is, first-degree holders who graduated from a UK university) and accounting for the degree class, university type and subject area of study.

Utilising the LFS data for the years 2002-2014, Longhi and Brynin (2017) employ decomposition techniques to explore ethnic differences in hourly wages. They find that Bangladeshi, Pakistani, and Black Caribbean male employees born in the UK experience pay gaps of 26%, 19% and 7%, respectively, relative to similarly qualified White British men. They illustrate that the younger age profile of ethnic minorities, their over-representation in low-paid occupations, their concentration in London and the level of education are key determinants of the wage gaps pertaining to observed characteristics. Similarly, Elliot and Lindley (2008) adopted decomposition methods to decipher the contribution of observed factors to the average weekly earnings gaps in the UK. They found that non-White natives face a raw log-pay penalty of 0.064 compared to White natives. Given that the component of earnings differential concerning the size of coefficients covers the largest part of this pay gap (that is, some observed characteristics are priced differently in the labour market between these two ethnic groups), the authors claimed that discrimination should be an element of the earnings differential.

The economic theory typically divides discrimination into taste-based, statistical, and unconscious (Guryan and Charles, 2013; Thijssen, 2016). Taste-based discrimination theory, which was proposed by Becker's (1971) models, suggests that certain employers have an antipathy for people from specific groups. This theory implies that some firms would choose to suffer economic costs (for example, by paying higher salaries for individuals that match their racial preferences) rather than employ members from ethnic minorities. Statistical discrimination is not associated with employers' prejudice, but it originates from the fact that profit-maximising firms may not have perfect information about the candidates' skills/profile and are reluctant to cover the cost associated with acquiring the missing information (Arrow, 1972; Phelps, 1972). Therefore, relying on previous statistical experience or *a priori* predominant sociological beliefs, those employers assess applicants' specific characteristics (such as their race) as a signal of the candidates' productivity, thus discriminating against ethnic minorities. The unconscious discrimination theory proposes that implicit interethnic attitudes result in discriminatory behaviours, even though the decision-makers strive to avoid stereotypes or racial preconceptions (Devine, 1989; Bertrand, Chugh and Mullainathan, 2005).

The above distinction of discrimination (especially the separation between the taste-based and statistical forms) is of major importance for policymakers, as policy response may differ according to the kind of discrimination. Hence, a notable amount of research examines the nature of discrimination in labour market outcomes by conducting experimental studies (see a systematic review of such studies in Neumark, 2018). Zschirnt and Ruedin's (2016) meta-analysis draws on data from 43 experimental works carried out in 18 countries (including the UK) to provide some evidence that statistical discrimination in the hiring process is lower for second-generation immigrants than those born outside the host country (that is, first-generation immigrants). This is possibly tied to the fact that more comprehensive information (such as the type and quality of qualifications) is available to employers for the former immigrant group. In contrast, the authors conclude that taste-based discrimination dominates, does not decrease for second-generation migrants and is more prominent for the most "visible" ethnic minorities. The data used in the present analysis do not directly capture these discrimination mechanisms. However, given that I concentrate on UK-born ethnic minority employees, many of whom are likely second- or higher-generation immigrants, I cautiously link any evidence of racial bias in wages to the taste-based form of discrimination (see subsection 4.5.2).

As mentioned earlier, field experiments address ethnic discrimination by almost exclusively centring on the recruitment process. Non-experimental approaches, such as the present, usually deal with wage inequalities or discrepancies in employment prospects. Thus, it remains mostly unknown how discriminatory attitudes in the hiring stages explain or carry over to wage outcomes at the market level (Neumark, 2018). One of the few exceptions is the work of Pager, Western, and Bonikowski (2009), who conducted a field experiment in New York City to describe a situation of employers "channelling" some ethnic minority applicants, during the interview, to inferior (and probably lower-paid) jobs of the organisational hierarchy (such as positions heavily reliant on manual tasks) than the initially advertised ones.

4.3 Data and methodology

To address the research questions, I exploit the Annual Population Survey (APS)[18] datasets pooled from 2013 to 2018, which collect information from the Labour Force Survey (LFS). The latter is an authoritative social survey, including comprehensive information about individual demographic characteristics, earnings, economic activity, education and training, health, and so forth. Hence, the nature and the coverage of this survey render it suitable for the current study.

The chosen sample comprises working-age employees (aged 19-65) who are born in the UK and all hold a first degree obtained from a British university. Information about wages is not available for self-employed individuals, who are, therefore, omitted from the present analysis. I have restricted the sample to the years mentioned above because the data regarding the university attended and subject area of first degree became available only after 2012 in the APS/LFS datasets. Furthermore, I exclude individuals who have also gained a higher degree (such as Masters or Doctorate) to eliminate potential omitted variable bias originating from self-selection to pursue higher types of qualification[19]. The sample also excludes employees who are still in full-time education and a few

[18] *"Office for National Statistics, Social Survey Division (2020). Annual Population Survey, 2004-2019: Secure Access. [data collection]. 15th Edition. UK Data Service. SN: 6721, http://doi.org/10.5255/UKDA-SN-6721-15." Due to data sensitivity, the analysis was remotely performed within the UK Data Service Secure Lab. The Statistical Disclosure Control required at least 10 individuals in each subgroup of the presented sample to make data available.*

[19] *As I will show in the robustness checks section (4.5.3), including postgraduate degree holders in the sample makes little difference to ethnicity's effect on wages.*

people who reported that their workplace is outside the UK. The total sample is composed of around 49,600 persons, with slight variations across variables due to differentials in missing cases.

4.3.1 Variables

The dependent variable I use in this work is the natural logarithm of the gross hourly wage in the employees' main job. I have converted wages to December 2018 constant prices (that is, real hourly wages) according to the Retail Price Index (RPI) released by the Office for National Statistics (ONS, 2020). To reduce the impact of outliers and possible reporting errors, I have trimmed the dependent variable by dropping the top and bottom 1% of the logarithm of wage distribution within men and women.

The key independent variable of interest is ethnicity, which classifies employees into four ethnic groups: White, Black, Asian, and Mixed/Other. Ethnic disparities in characteristics are likely to be present within the minority groups that define these major categories. For example, as Blackaby et al. (1999) mentioned, the Indian ethnic group differs markedly from the Pakistani/Bangladeshi community in terms of assimilation, economic isolation, and diverse levels of "enclavement" (that is, the degree of clustering of ethnic minorities in local geographical areas or neighbourhoods). However, data constraints arising from the relatively small number of observations for non-White employees prevent a more detailed disaggregation. Therefore, I implement the broad classification of ethnicity proposed by the 2011 National Statistics, except that I group the "Other" and "Mixed" ethnic minorities into one category (due to the limited number of employees within the "Other" ethnic group).

To explore ethnic variations in wages, I account for many covariates that influence the level of earnings, based on the relevant research studies discussed in section 4.2. I briefly present the main independent variables below, while Table 4.A1 in the Appendix provides a detailed description of them.

Demographic characteristics: This group of covariates comprises age, age-squared, the region of workplace (which in most cases coincides with the region of residence), marital status (a binary variable representing whether an employee is married/in a civil partnership or not), an indicator of the existence of any children aged under 19 in the family, and a dummy variable for chronic health problems (that is, illnesses and other conditions that persist for over one year). In the absence of information on employees' experience in administrative datasets, a common approach in the literature is to approximate it by using age. This practice usually works well in empirical research[20] (e.g., Carmichael and Woods, 2000), especially for men, as they are not equally affected by factors related to family formation and commitment to workforce careers (Lindley, Dale and Dex, 2004). Moreover, I incorporate the quadratic term of age in the analysis to capture the diminishing returns of age in earnings.

Higher education characteristics: This set of independent variables contains the subject area and class of first degree and the type of university attended. Following the approach of Chowdry et al. (2013), I split the higher education institutions into “high status” and “other” universities. The high-status universities comprise 41 UK institutions, including the well-regarded Russell Group (RG) universities, plus any HE institutions demonstrating a higher ranking than the lowest RG university, based on the 2014 Research Excellence Framework (REF). I have grouped the subject of study of the first degree into seven categories: Health, Sciences, Engineering/Technology, Social studies, Law/Business/Finance, Arts/Humanities/ Education, and Combined subjects. The latter category covers degrees in one or over one subject area (for example, BSc in Economics & Mathematics). Furthermore, I aggregate the class of degree into a “good degree” dummy variable, which denotes whether or not an employee holds a first-class or an upper second-class degree.

Occupation-Sector characteristics: These factors refer to the occupation of employees (categorised in “managerial/professional jobs” and “other occupations”), tenure (that is, the years worked for the current employer), the industry sector (clustered in six broad categories), and the workplace size (classified into “micro/small” enterprises with less than 50 employees and “medium/large” companies with over 50 workers). This analysis also takes into account information about the weekly total working hours (including overtime), which measure work intensity, as well as dummy variables designating whether working on a part-time/full-time basis, on a permanent basis (or on a fixed-time contract), and whether an employee holds a position in the public sector.

---

[20] Some studies approximate experience by using current age minus age that a person completed full-time education. However, apart from the fact that this approach has several shortcomings (Shields and Price, 1998), the variable capturing the graduation year exhibits a large percentage of missing (not reported) cases in the APS datasets.

4.3.2 Descriptive evidence

One possible explanation for ethnic gaps in wages is that ethnic minority employees have distinct personal characteristics from White people. In this subsection, I descriptively explore the key determinants of earnings and present the raw wage differential between ethnic groups. The total proportion of non-White employees amongst UK-born graduates is relatively small, representing 6.0% of the sample for women and 5.8% for men (see Table 4.1). Workers from Indian, Chinese and "any other Asian" backgrounds constitute the highest-paid ethnic minorities, earning, on average, more than their White counterparts. Notwithstanding, all other ethnic groups see lower average wages than the White majority group, and the ethnic gaps are substantially larger for Pakistani and Bangladeshi employees. For example, an average Bangladeshi UK-born graduate earns nearly four pounds per hour less (£15.03) than a White worker (£18.93) in the UK labour market.

Tables 4.A2 and 4.A3 in the Appendix present the sample means[21] of the variables used in this paper by White status[22] and gender, respectively. The raw non-White–White hourly earnings difference stands at 0.06 log units (in favour of White employees), and this gap is statistically significant at the 1% level. Based on the probability distribution of log (wage) depicted in Figures 4.1 and 4.2, the average real hourly earnings are generally lower for ethnic minorities across both genders, but the pay disadvantage is more prominent amongst men. Looking at the tails of the wage distribution, the proportion of highly paid Black men is significantly smaller than that of their White counterparts, suggesting that Black male workers are under-represented in the high-salaried occupations and sectors. The fact that most non-White people are concentrated in London (37.9%, Table 4.A2 in the Appendix) confirms the historical pattern described by official statistics and the migration literature (Phillips, 1998; Craig, 2012). Moreover, a significantly higher proportion of White people (72%) are married or cohabiting than non-White employees (54%), albeit the latter are more likely to have dependent children.

[21] *All figures presented in the descriptive analysis are unweighted and should not be considered as population means. Each yearly APS dataset provides weights (based on the respondents' three main characteristics: gender, location, and age) to account for the sample design and non-response bias. However, when merging APS datasets for multiple years to improve the sample size, the individual-weight variables (which are calculated on an annual basis) cannot be used.*

[22] *The data providers' restrictions regarding the underlying cell sizes prohibit display of the detailed distribution of the variables by ethnic groups within each gender.*

The present work primarily analyses the life-course wage outcomes of UK-born graduates, rather than differences existing at a specific point in time of their career (Walker and Zhu (2013) adopted a similar approach). The logic behind this is that solely focusing on labour market outcomes at the early stages of working life (where the wage dispersion is limited) might bias the general picture, given that it usually takes a few years for young graduates to find their job match.

Overall, the age profile of employees from a non-White ethnic background is younger (34 years old, on average) than that of White people (39 years). Figure 4.3 reveals that earnings increase drastically until the early 40s for all four ethnic groups of employees. Until the end of their 20s, ethnic minorities exhibit higher wages (on average) than White employees (except for Black men), although the low number of observations within each age band compromises the statistical significance of wage differentials. For example, the average log hourly wage for Asian men aged between 26-30 years is 2.77, whereas the corresponding figure for White men stands at 2.68. This picture is reversed for older male employees, as White graduates aged over 40 years earn, on average, more than the rest of ethnic groups. I discuss these differences in detail and provide possible explanations in subsection 4.4.2.2. Specifically, I impose age restrictions in the regression analysis to show that early-career employees from ethnic minority backgrounds (aged below 30) face substantially lower or no ethnic pay gaps than older employees of similar observed characteristics.

As Table 4.A4 in the Appendix illustrates, raw hourly wages differ in a statistically significant way (in favour of White employees) within almost all subgroups of demographics, higher education, and job-relevant characteristics, indicating a multi-dimensional nature of ethnic penalties. The following tables show that ethnic penalties are more pronounced amongst men than women. This may be partially attributed to the well-established gender income inequality (Machin and Puhani, 2003; Longhi and Platt, 2008) and discrepancies in occupational choices between females and males (Manning and Swaffield, 2008). In the results section of this paper (decomposition analysis), I highlight the contribution of each main group of observable covariates to the earnings differential between the two major ethnic groups (White and non-White employees).

The labour market returns vary significantly across subjects when the first-degree major is considered (Table 4.2). For example, Engineering and

Technologies subjects yield wages averaging 0.16 and 0.37 log points higher than the Arts/Humanities/Education category, for women and men, respectively. It is well-known from the UK literature that STEM and LEM subjects offer a higher premium relative to other degrees (Conlon and Patrignani, 2011; Walker and Zhu, 2011, 2013; Britton et al., 2016; Walker and Zhu, 2018; Belfield et al., 2018a, 2018b). Across all subject areas, White men earn more than their Black counterparts. The average White–Black wage gap ranges from 0.07 log points for male holders of Arts/Humanities/Education degrees to 0.29 log points for the combined subjects. In contrast, there is a more mixed picture in wage differences between the rest of ethnic groups, especially amongst women. Overall, the data shows that in most degree subject areas (except for Health-related degrees), the offered returns (in terms of average wages) remain lower for non-White minorities, and these differences are statistically significant at the 1% or 5% levels (Table 4.A4 in the Appendix).

The proportion of good degree holders is comparable between both broad ethnic groups, standing at 56.2% for non-White and 57.1% for White workers. However, there is an average wage penalty of 0.03 log points faced by ethnic minorities graduating with a good degree (Table 4.A4 in the Appendix). The pay gap is three times bigger for graduates with a lower class of degree. As in the findings mentioned earlier, the differences in wages are greater amongst men, particularly in the case of the Black minority (Table 4.3). Feng and Graetz (2017) establish causal effects of the class of degree on earnings. Naylor et al. (2016) point out that the wage premium connected with a good degree increased over time in the 1990s along with the rise in higher education (HE) participation rates.

Interestingly, in line with the HE expansion, the percentage of employees graduating with a first or an upper second class of degree enlarged dramatically, from around 50% for cohorts graduated in the early 1990s to over 75% in 2018 (Figure 4.4). Controlling for the degree class in the following regression analyses likely captures differences in the underlying (unobserved) ability. Nevertheless, this is an imperfect measure of ability given that degree classes may not be equivalent across institutions. As the preceding studies mention, degree class could also act as a signal of productivity to employers, especially at the early career stages.

Substantial differences in raw wages between ethnic groups also exist within broad types of university attended (Figure 4.5). The earnings differences are generally higher for ethnic minority men who graduated from a non-high-status university than a prestigious institution. This is likely linked to the fact that many elite universities provide students from disadvantaged backgrounds with application support and professional opportunities with partner employers, thus strengthening their career skills and networks (Russell Group, 2019). However, the pay disadvantages for Black male employees remain large within both university types. The percentage of non-White minorities who attended a high-status institution (26.2%) is significantly lower than that of White employees (33.5%). This is consistent with a considerable body of research which argues that specific ethnic minorities (notably those from Black, Pakistani, and Bangladeshi backgrounds) are under-represented in the "old" (pre-1992) and Russell Group universities (Coffield and Vignoles, 1997; Shiner and Modood, 2002; Boliver, 2013, 2016). Given that prestigious institutions select high-quality students, who, in turn, experience better salaries in the labour market (Dale and Krueger, 2014; Walker and Zhu, 2018), controlling for the university type in the econometric analysis (on top of the subject of study and degree class) helps explain a significant component of the wage differential between White and ethnic minority graduates.

Although the chosen sample comprises employees with comparable human capital (that is, university graduates), the occupational distribution varies significantly between the ethnic groups. More specifically, 76.2% of White employees have a managerial/professional job compared to 69.9% among non-White minorities (Table 4.A2 in the Appendix). Some studies have examined the disparities in occupational segregation by gender and ethnic groups (Cohen and Huffman, 2007; Manning and Swaffield, 2008; Li and Heath, 2010; Lindley, 2016; Mok and Platt, 2018), stating that much of the earnings difference is because of the concentration of ethnic minorities in low-paid occupations. Clark and Drinkwater (2002) found a negative correlation between the UK areas with a high concentration of ethnic minorities and their proportion in the professional/managerial occupations. Table 4.4 presents the wage distribution based on a more detailed classification of occupations by gender and ethnic groups. In line with the previous findings, the wage gaps favouring White workers are larger amongst men and more distinct within the well-compensated jobs. For example, Black men occupying managerial or professional positions experience an average pay disadvantage of 0.18 log points compared to White men (Figure 4.6).

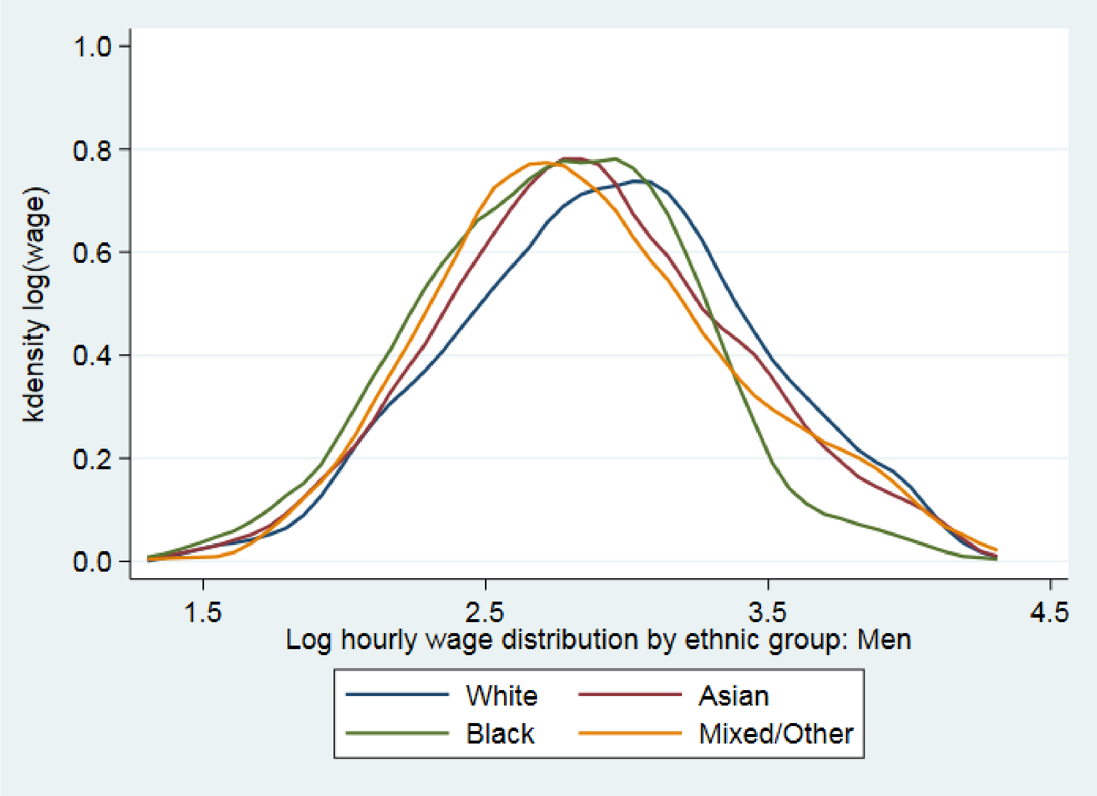
Most non-White employees (36.4%) are concentrated in the Public Admin/Education/Health sectors, followed by the financial sector (27.5%). Figure 4.7 shows the average level of wages across major industry sectors. On average, the best-paid sectors for non-White workers are the Manufacturing/Construction and Transport/Communication industries (Table 4.A4 in the Appendix), but, again, within the former category, ethnic minorities are under-represented compared to White employees (7.1% versus 10.8%). Although most people from non-White ethnic backgrounds (65.4%) work in medium/large companies, they are still penalised by 0.06 log points (on average) in the labour market, while the corresponding pay gap is slightly higher within the micro/small enterprises (0.08 log points). Finally, the average number of years in current employer (job tenure) is significantly lower for ethnic minority employees (5.1 years) than their White counterparts (7.8 years, see Table 4.A2 in the Appendix). As I will show in section 4.4, this variable is positively associated with wages, and the difference between White and ethnic minority employees in firm tenure contributes to the size of ethnic pay penalties.

Table 4.1. Frequencies and real hourly wage: by ethnicity and gender

Ethnicity	**Frequencies**				**Mean real hourly wage (£)**		
	Women	**Men**	**Total**	**Total %**	**Women**	**Men**	**Total**
White	25,624	21,111	46,735	94.1%	16.77	21.54	18.93
Indian	468	439	907	1.8%	17.44	21.22	19.27
Black/African/Caribbean/Black British	350	188	538	1.1%	16.41	17.39	16.76
Pakistani	240	246	486	1.0%	13.75	17.63	15.71
Bangladeshi	68	57	125	0.3%	14.36	15.83	15.03
Chinese	68	63	131	0.3%	17.03	21.88	19.36
Any other Asian background	58	44	102	0.2%	16.74	25.31	20.43
Mixed/Multiple ethnic groups	293	212	505	1.0%	15.41	19.54	17.14
Other ethnic group	80	47	127	0.3%	16.53	21.40	18.33
Total	**27,249**	**22,407**	**49,656**	**100%**	**16.73**	**21.43**	**18.85**

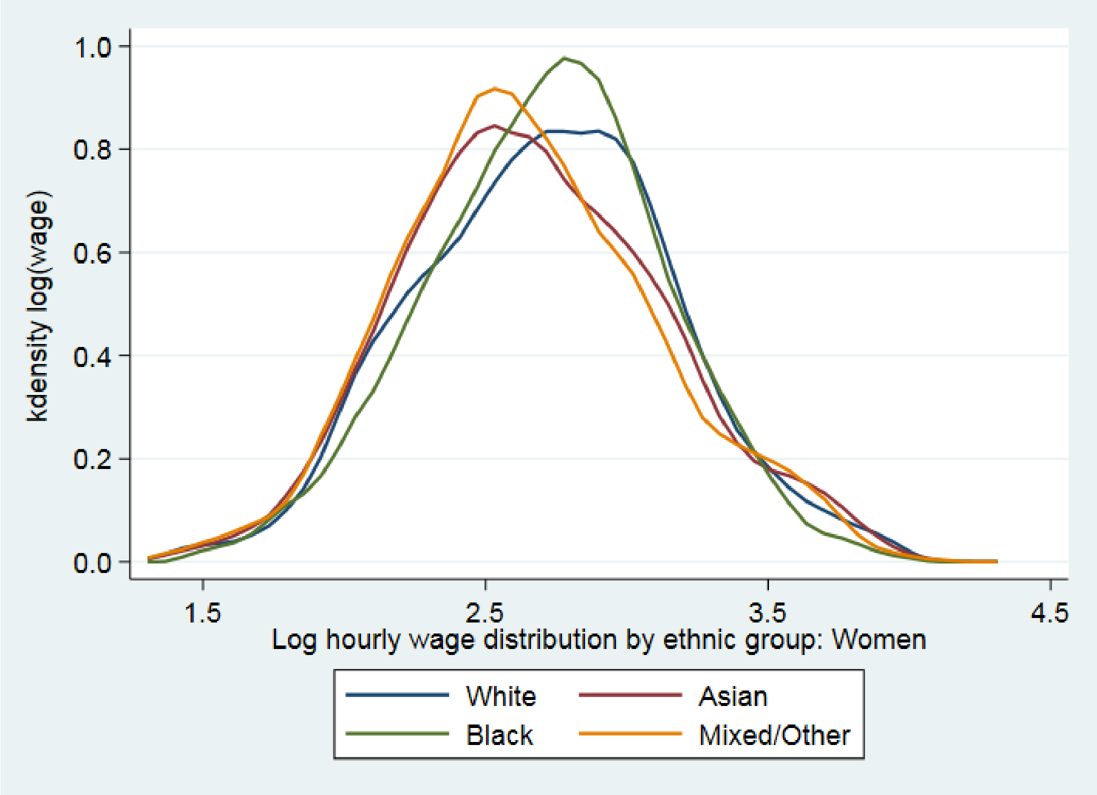
Source: APS 2013-2018, author's own calculations

Figure 4.1. Probability distribution (Kernel density) of log(wage): by ethnic group – Men



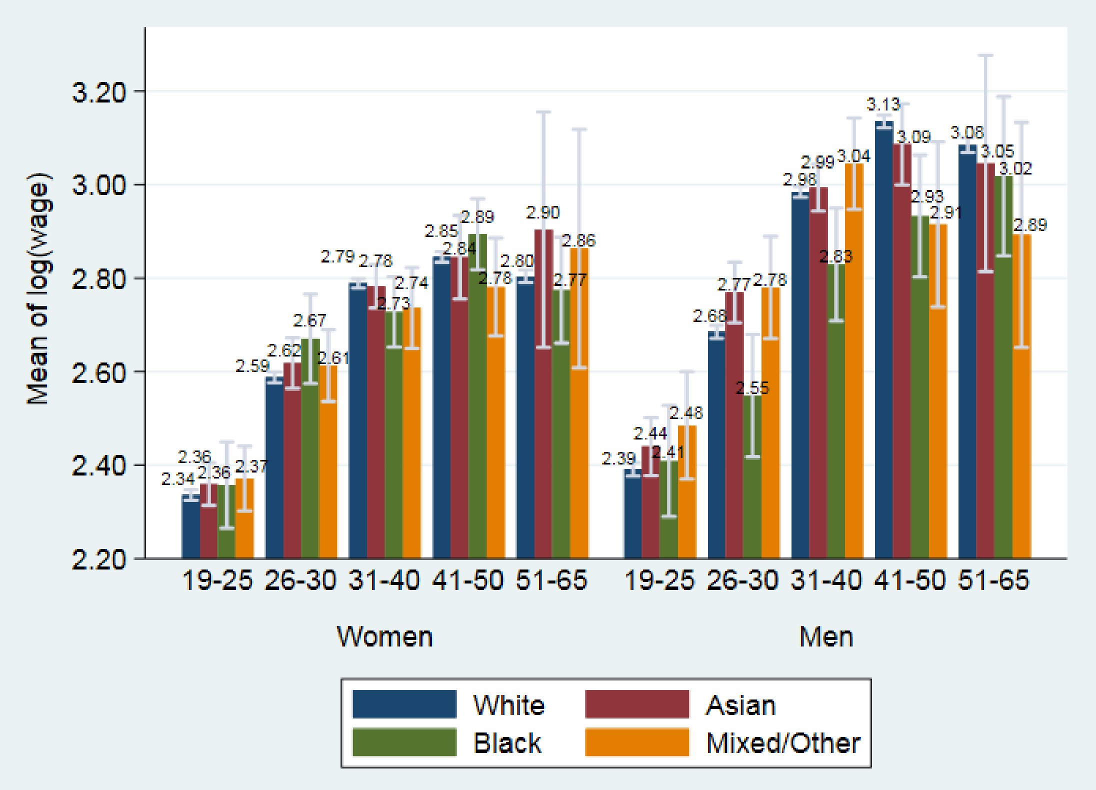
Source: APS 2013-2018

Figure 4.2. Probability distribution (Kernel density) of log(wage): by ethnic group – Women



Source: APS 2013-2018

Figure 4.3. Mean log(wage): by age bands, gender, and ethnic group



Note: The error bars represent the 95% confidence interval of the mean log hourly wage of each ethnic group.
Source: APS 2013-2018

Table 4.2. Mean log (wage): by subject area of first degree, gender, and ethnicity

Subject area of first degree	Women					Men				
	White	Asian	Black	Mixed/ Other	Total	White	Asian	Black	Mixed/ Other	Total
Health	2.80	2.82	2.77	2.81	2.80	2.93	3.01	*	2.90	2.93
Sciences	2.69	2.72	2.76	2.53	2.69	2.92	2.85	2.74	2.80	2.91
Engineering/Technology	2.82	2.68	*	*	2.81	3.10	2.89	2.84	3.10	3.09
Social studies	2.69	2.62	2.73	2.58	2.68	2.95	3.03	2.78	3.00	2.95
Law/Business/Finance	2.76	2.63	2.77	2.68	2.75	2.97	2.79	2.85	2.96	2.95
Arts/Humanities/Education	2.65	2.59	2.63	2.56	2.65	2.73	2.75	2.66	2.61	2.72
Combined subject	2.73	2.73	2.72	2.74	2.73	2.96	2.87	2.67	2.76	2.95
Total	**2.71**	**2.67**	**2.72**	**2.65**	**2.71**	**2.93**	**2.86**	**2.75**	**2.85**	**2.93**

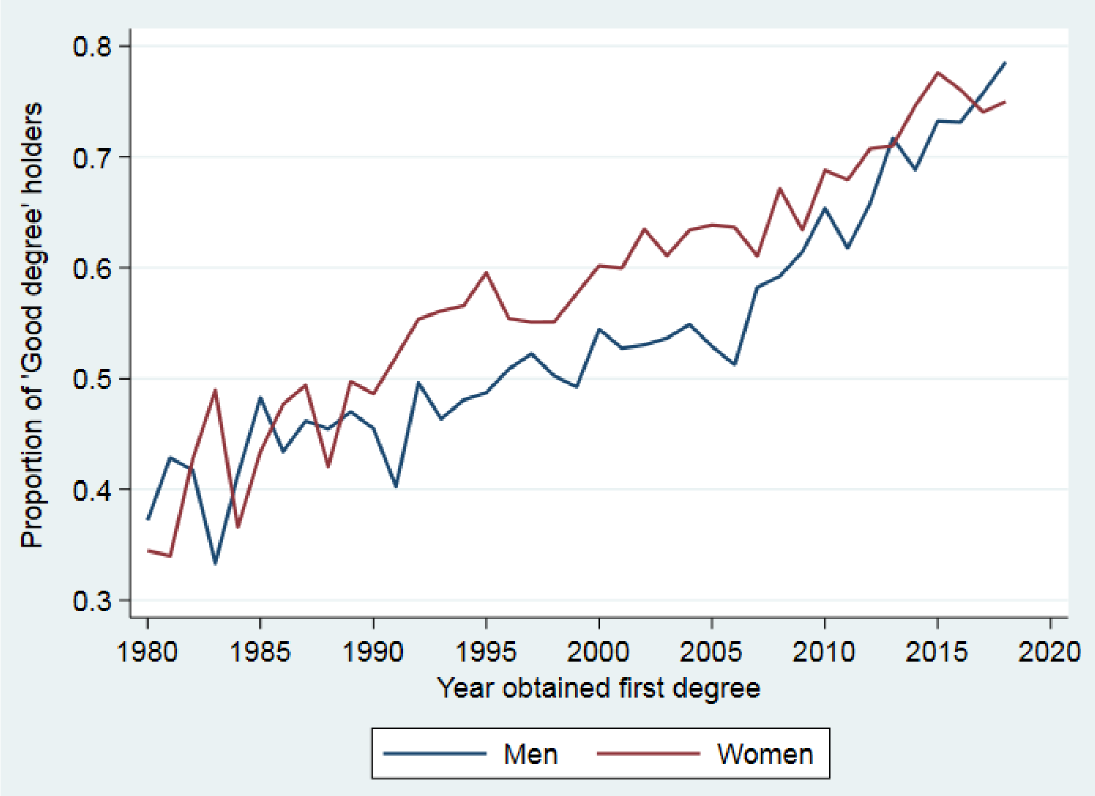
*Note: * denotes figures that are not displayed due to the small size of the underlying cells (<10 observations).*
Source: APS 2013-2018

Table 4.3. Mean log (wage): by class of first degree, gender, and ethnicity

Class of first degree		Women					Men				
		White	Asian	Black	Mixed/ Other	Total	White	Asian	Black	Mixed/ Other	Total
First	Log(wage)	2.73	2.75	2.70	2.54	2.72	2.95	3.00	2.70	2.85	2.95
	Observations	2,969	65	23	41	3,098	2,219	73	12	25	2,329
Upper second	Log(wage)	2.72	2.70	2.73	2.68	2.72	2.94	2.91	2.79	2.92	2.94
	Observations	12,441	459	164	180	13,244	9,033	390	85	123	9,631
Lower second	Log(wage)	2.67	2.61	2.69	2.58	2.67	2.92	2.80	2.68	2.73	2.91
	Observations	6,271	283	112	102	6,768	6,231	265	65	72	6,633
Third	Log(wage)	2.64	2.61	2.69	2.60	2.64	2.89	2.81	2.69	2.76	2.89
	Observations	640	27	21	16	704	1,091	38	10	11	1,150
Pass	Log(wage)	2.77	2.75	2.84	2.89	2.77	2.94	2.85	*	2.93	2.94
	Observations	1,953	28	13	15	2,009	1,488	35	*	13	1,536
Other/Unknown	Log(wage)	2.76	2.63	2.73	2.75	2.76	2.90	2.74	*	2.88	2.89
	Observations	1,331	38	16	18	1,403	1,027	47	*	15	1,100
Total	**Log(wage)**	**2.71**	**2.67**	**2.72**	**2.65**	**2.71**	**2.93**	**2.86**	**2.75**	**2.85**	**2.93**
	Observations	**25,605**	**900**	**349**	**372**	**27,226**	**21,089**	**848**	**172**	**259**	**22,368**

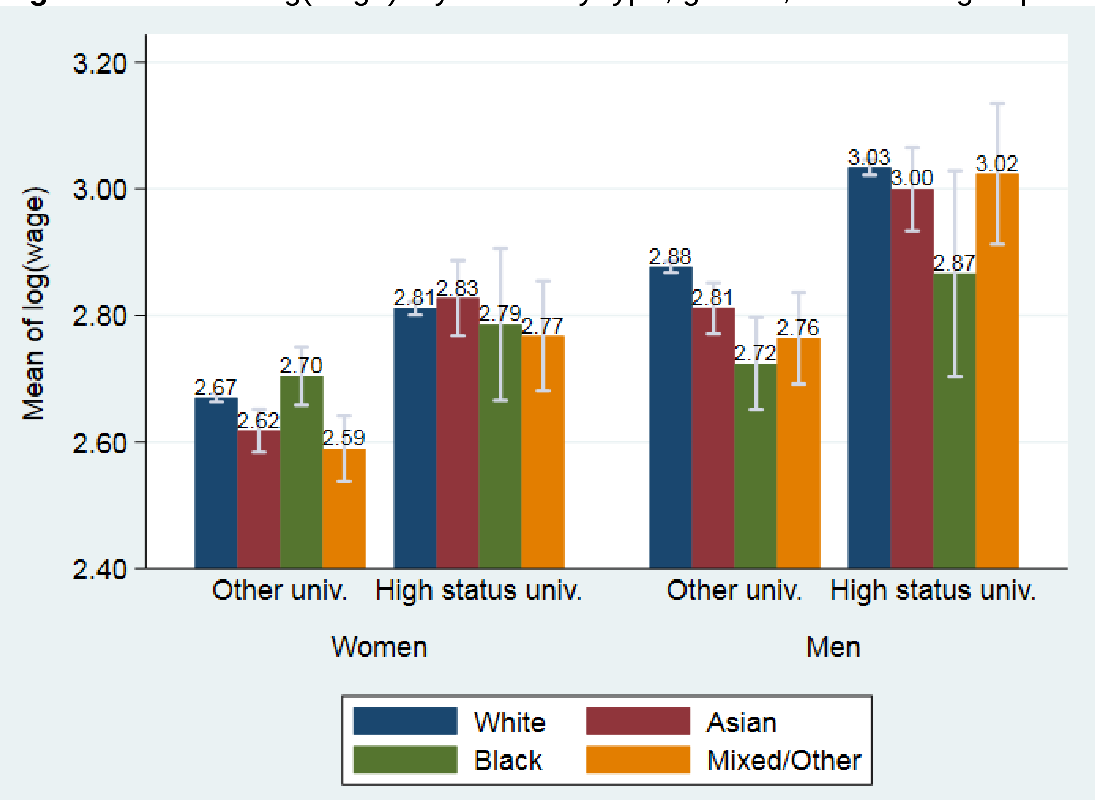
*Note: * denotes figures that are not displayed due to the small size of the underlying cells (<10 observations). Totals have been adjusted accordingly.*
Source: APS 2013-2018

Figure 4.4. Proportion of "good degree" holders by gender and year obtained first degree (1980-2018)



Source: APS 2013-2018

Figure 4.5. Mean log(wage): by university type, gender, and ethnic group



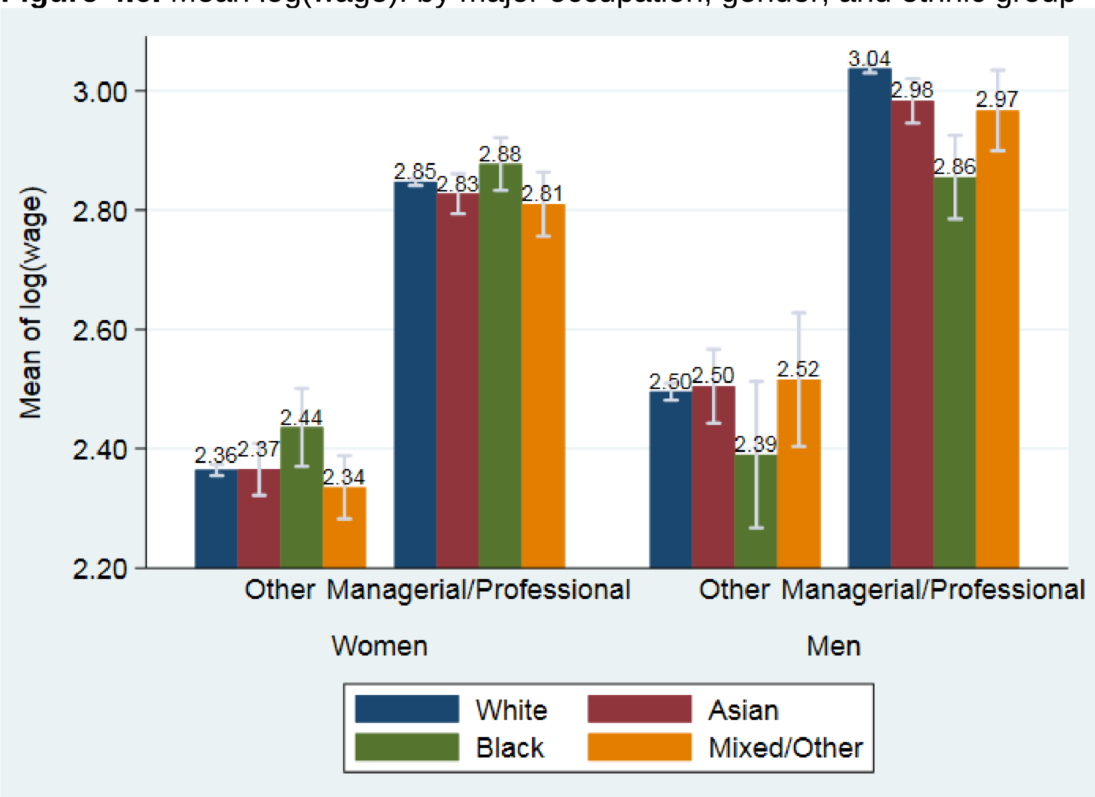
Note: The error bars represent the 95% confidence interval of the mean log hourly wage of each ethnic group.
Source: APS 2013-2018

Table 4.4. Mean log (wage): by occupation group and ethnicity

Occupation group		White	Asian	Black	Mixed/ Other	Total
"Managers, Directors & Senior Officials"	Log(wage)	3.10	3.01	2.93	3.04	3.09
	Observations	5,783	155	41	53	6,032
"Professional Occupations"	Log(wage)	2.95	2.97	2.93	2.89	2.95
	Observations	19,653	627	203	230	20,713
"Associate Professional & Technical Occupations"	Log(wage)	2.82	2.79	2.75	2.80	2.81
	Observations	10,179	456	123	153	10,911
"Administrative & Secretarial Occupations"	Log(wage)	2.53	2.58	2.57	2.46	2.53
	Observations	4,585	238	79	85	4,987
"Skilled Trades Occupations"	Log(wage)	2.65	2.62	*	*	2.65
	Observations	784	14	*	*	798
"Caring, Leisure & Other Service Occupations"	Log(wage)	2.26	2.28	2.33	2.35	2.27
	Observations	2,333	70	34	31	2,468
"Sales & Customer Service Occupations"	Log(wage)	2.34	2.27	2.32	2.32	2.34
	Observations	2,026	151	34	41	2,252
"Process, Plant & Machine Operatives"	Log(wage)	2.48	2.45	*	*	2.48
	Observations	389	10	*	*	399
"Elementary Occupations"	Log(wage)	2.16	2.19	2.19	2.24	2.17
	Observations	983	29	17	23	1,052
Total	**Log(wage)**	**2.81**	**2.77**	**2.73**	**2.73**	**2.81**
	Observations	**46,715**	**1,750**	**531**	**616**	**49,612**

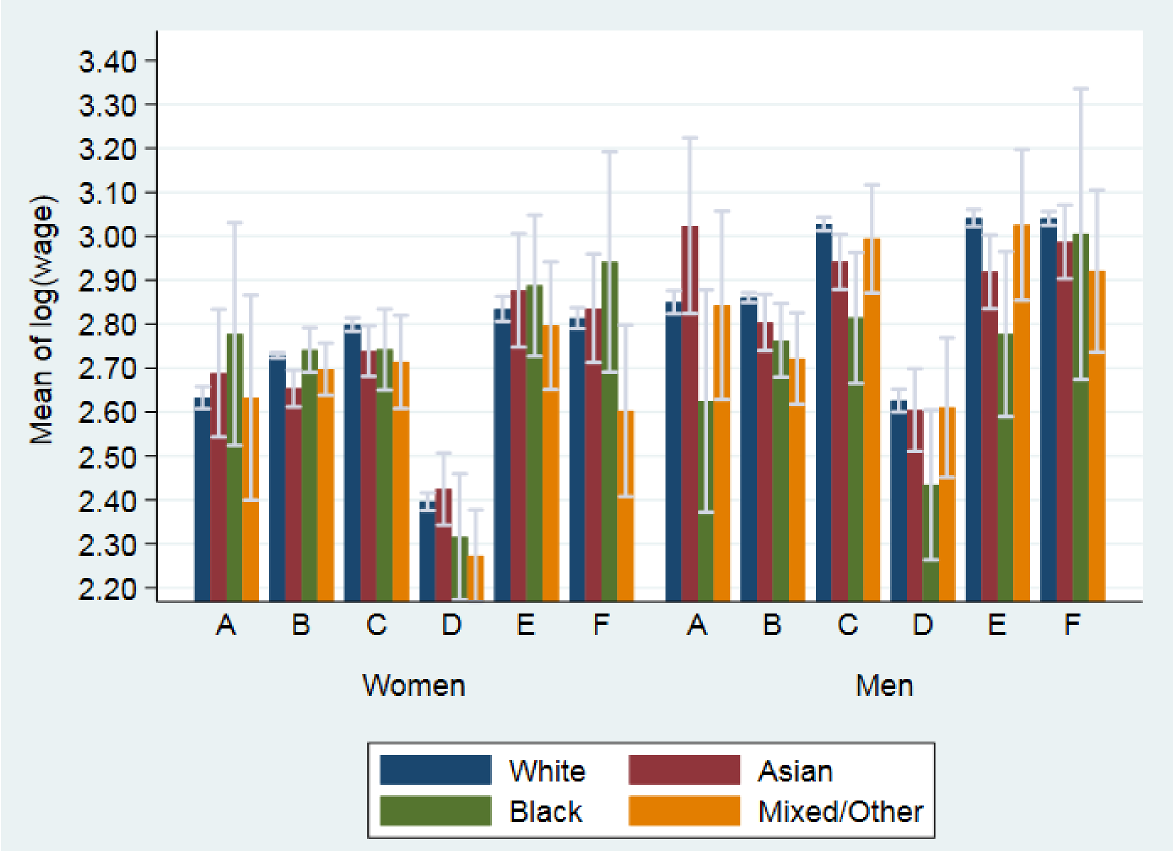
*Note: * denotes figures that are not displayed due to the small size of the underlying cells (<10 observations). Totals have been adjusted accordingly.*
Source: APS 2013-2018

Figure 4.6. Mean log(wage): by major occupation, gender, and ethnic group



Note: The error bars represent the 95% confidence interval of the mean log hourly wage of each ethnic group.
Source: APS 2013-2018

Figure 4.7. Mean log(wage): by major industry, gender, and ethnic group



Note: The error bars represent the 95% confidence interval of the mean log hourly wage of each ethnic group. A: Other industry (e.g., "agriculture/forestry/fishing", "energy and water", "other services"); B: Public Admin/Education/Health; C:Banking/Finance; D:Trade/Hotels/Restaurants; E:Transport/Communication; F:Manufacturing/Construction
Source: APS 2013-2018

4.3.3 Econometric framework

4.3.3.1 OLS model specification

I extend the traditional Mincer's (1974) earnings equation by accounting for many human capital, occupation and personal/family factors alluded to in the preceding section. In particular, the preferred model specification employed in the Ordinary Least Squares (OLS) analyses is:

$$\ln(W_i) = \mathbf{x'}_i\boldsymbol{\beta} + \theta Ethnic_i + \ \varepsilon_i \qquad (1)$$

where W_i represents the real gross hourly wage of employee i; $\mathbf{x_i}$ denotes the vector of individual-level covariates; $\boldsymbol{\beta}$ are the respective coefficients of the vector and ε_i is an independently distributed random error, which captures all other unobserved determinants of earnings. The coefficient of the categorical variable $Ethnic$ (θ) is the parameter of interest, as it measures the average "ethnic penalty" (*ceteris paribus*) in wages of people from ethnic minority backgrounds (Black, Asian, and Mixed/Other). The vector $\mathbf{x}$ contains all the control variables described in the previous section: demographic traits (age, age squared, region of workplace, partnered, any dependent children in family aged under 19, and health problems lasting over one year), higher education characteristics (subject area of first degree, high-status universities, and good degree), and job-pertinent factors (managerial/professional occupation, industry sector, medium/large enterprises (>=50 employees), public sector, part-time work, permanent job, years in current employer (tenure), and total usual weekly hours in the main job). It also includes dummies for the year and month of the survey (cohort fixed effects) to account for any systematic differences in the respondents' characteristics deriving from the survey design.

In the main regression analysis, I present three modelling stages to demonstrate how including specific characteristics changes each ethnic minority group's coefficient. The basic idea is to investigate whether the extent of the relationship between ethnicity and the outcome of interest (log hourly wage) is altered when I incorporate additional factors in the models. First, I include only demographic variables (region of the workplace, age, age squared, family and health characteristics) and survey dummies. Subsequently, I sequentially add in a wide range of variables related to higher education (second stage) and occupation/sector characteristics (third stage). This procedure enables me to detect whether wage gaps between White and ethnic minorities decrease (or increase, depending on the direction of the omitted variable bias) once I control for each of those groups of wage determinants.

Throughout the analysis, I present gender-specific results to account for the fact that some characteristics are valued differently in the labour market between men and women, while the impact of unobserved wage determinants and participation patterns may also vary across genders.

4.3.3.2 Quantile regression

I also explore heterogeneous effects on real hourly earnings by running quantile regressions (QR), which calculate conditional quantiles of the wage distribution as defined by the linear specification (Koenker and Bassett, 1978). In doing so, I provide a more thorough picture of the impact of ethnicity on earnings. In contrast to the OLS estimates that assume fixed effects across the wage distribution, QR allow for unequal effects of the regressors along the earnings distribution spectrum. Specifically, the QR method permits calculating the effect of ethnicity at different points of the conditional earnings distribution, not simply its conditional mean (or median). Unlike the OLS method, which produces the $\boldsymbol{\beta}$ coefficients by minimising the quantity $\sum_{i=1}^{n} e_i^2$ (sum of squared residuals), QR estimates are based on linear programming (LP) techniques. Within the LP framework, the QR method relies on defining the q_{th} quantile and estimating the coefficients β_q by minimising the following function:

$$Q(\beta_q) = \sum_{i:\, y_i \geq \mathrm{x}_i'\beta}^{N} q\left|y_i - \mathrm{x}_i'\beta_q\right| + \sum_{i:\, y_i < \mathrm{x}_i'\beta}^{N} (1-q)\left|y_i - \mathrm{x}_i'\beta_q\right| =$$

$$= \sum_{i:\, \varepsilon_i \geq 0}^{N} q|\varepsilon_i| + \sum_{i:\, \varepsilon_i < 0}^{N} (1-q)|\varepsilon_i| \qquad (2)$$

where $q \in (0,1)$. If $q = 0.5$ (for the median), then this becomes a least-absolute-deviations (LAD) optimisation problem, as symmetric weights are assigned to all observations. In all other cases ($q \neq 0.5$), the QR method allocates asymmetric weights ("penalties") for overestimation $((1-q))$ or underestimation (q) of the residuals ε_i (for more formal details and applications of QR see Koenker, 2005). The QR approach is more robust than the OLS method to outlying observations and errors that do not follow the normal distribution (Baum, 2016).

4.3.3.3 Decomposition method

Many labour market studies (e.g., Machin and Puhani, 2003; Longhi, Nicoletti and Platt, 2013; Christopoulou and Monastiriotis, 2014) use decomposition methods to examine mean wage differences on a counterfactual (out-of-sample) basis. Decomposition approaches divide the wage gap into explained and unexplained components, aiming to identify each part's relevant contribution to the total wage differential. In this paper, I implement a decomposition technique proposed by Neumark (1988), which is alternative to the standard Blinder-

Oaxaca (1973) procedure. The method used here was further developed by Oaxaca and Ransom (1994), who allowed the interpretation of results to be generalised at the market level. This decomposition approach separates the wage differences between White and non-White employees ($\Delta\bar{W} = \ln(\bar{W}_{i,White}) - \ln(\bar{W}_{i,nonWhite})$) into two components, as follows:

$$\Delta\bar{W} = \{(\bar{X}_{i,White} - \bar{X}_{i,nonWhite}) * \beta_{pooled}\} + \\ + \{(\beta_{White} - \beta_{pooled}) * \bar{X}_{i,White} + (\beta_{pooled} - \beta_{nonWhite}) * \bar{X}_{i,nonWhite}\} \quad (3)$$

The first component in brackets (known as “endowments effect” or explained part) captures the portion of the wage differential between White and non-White workers relating to differences in average employees’ characteristics ($\bar{X}_i$), evaluated at an “average” term denoted by the β_{pooled} coefficient, which is an estimate from a pooled OLS regression on both White and non-White groups. In other words, the first term reflects the average change in non-White employees’ wages if they had the White workers’ predictor levels (that is, if both ethnic groups had similar observed characteristics), weighted by the reference coefficient, β_{pooled}.

The second term in brackets (known as the unexplained part) describes the earnings gap linked to differences in the coefficients. It represents the earnings differential deriving from discrimination and other unobserved characteristics between White and non-White workers. This term is divided into two sub-components: a) the left-hand side one shows the White employees’ advantage (which reflects a “positive discrimination” or an overvaluation of the White ethnic group), multiplied by the White characteristics mix; and b) the right-hand side one represents the non-White employees’ disadvantage (which reflects “negative discrimination” or an undervaluation of the non-White minorities), weighted by the average non-White characteristics mix.

The reference coefficient β_{pooled} is often considered a non-discriminatory parameter. Specifically, Neumark (1988) assumes that there is a set of observable traits that explain any productivity discrepancies between two population groups, and the unexplained differences should exclusively reflect discrimination. However, this strong assumption is unlikely to hold in most empirical studies (Jann, 2008; Elder, Goddeeris and Haider, 2010). Moreover, I adopt Jann's (2008) suggestion to include the treatment dummy of "non-White" in the pooled regression as an additional explanatory variable to control for the fact that Neumark's method may incorrectly transmit a proportion of the unexplained component of the wage gap into the explained part.

Finally, I further split the explained part of wage differential into three sub-components, namely "demographic", "higher education", and "occupation/sector" characteristics, each of which contains the independent variables described in subsection 4.3.1. This detailed decomposition allows me to disentangle the average contribution of each group of covariates to the earnings gap between White and non-White employees, and it could provide valuable insights to policymakers aiming to suppress ethnicity barriers in labour market outcomes.

4.3.3.4 Caveats

The sample used in this analysis comprises only employees, thus excluding self-employed persons, working-age students in full-time education, retirees, unemployed people, or those who work but do not disclose wages. The decision to take part in the labour market as an employee is likely pertinent to one's optimising behaviour with regards to his/her accumulated human capital (Reimers, 1983). Omitted variables presumably affect the individual productivity of employees differently compared, for example, to the self-employed sector, thus systematically biasing the OLS coefficient estimates of the wage function. In the current setting, the non-random nature of the sample may cause a correlation of the omitted (unobserved) wage determinants with the probability of participating in the wage sector. Additionally, the probability of working in salaried employment may also be correlated with the regressors (particularly with ethnicity, which is the key independent variable of interest). As a corollary, if both these assumptions are correct, the observed pay gaps between ethnic groups estimated using OLS would overstate or understate the real population impact of ethnicity on wages. The situation described above, which implies a non-random inclusion into paid employment, is termed sample selectivity bias.

Neal (2004) argued that, because of their lower total family income, Black women have, on average, a higher motivation to participate in the labour market than White women of comparable educational levels. He posited that this partially explains why the wage gap between Black and White employees is historically lower amongst women than men in the US. According to this view, when assuming no racial discrimination, the underestimation of the female Black-White wage gap (because of the sample selection bias) might follow from two mechanisms. First, Black women may have different incentives to enter the labour market than White females because of unequal spousal income and marriage patterns. Second, those Black women who participate in the wage sector should have some productivity characteristics (such as ability or economic motivation) that employers value higher compared to the Black females who do not work. In addition, some ethnic minority people (particularly those from Pakistani and Bangladeshi backgrounds) are more likely to be self-employed than the White British group (Clark, 2015). Therefore, the unobserved characteristics may also include the impact of spending a period in self-employment on current wages (if self-employed people subsequently move to paid employment), which is not captured by the data.

In the present analysis, I do not adjust the model estimates for the sample bias originating from the selection into paid employment. Heckman (1979) showed that using a probit specification to predict the probability of participating in the wage sample and, subsequently, incorporating the inverse Mill's ratio estimated from the reduced form of probit model into the earnings function lead to consistent estimates of the wage equation. However, Heckman's two-stage method requires a valid instrumental variable to satisfy the exclusion restriction, thus affecting the probability of participating in the wage sector but not directly influencing earnings. The literature often arbitrarily adopts characteristics relating to the family background (such as family income) as candidate instruments. In the absence of convincing exogenous variables in the APS datasets, I opt to not correct the coefficients for sample selection bias because if the chosen instruments fail to meet the necessary conditions, the adjusted estimates may be less reliable than the OLS ones (Wolfolds and Siegel, 2019). Besides, among men, it is highly possible that sample selectivity does not play a substantial role in explaining the pay differential between ethnic groups. For example, Blackaby et al. (2002), using the LFS data, illustrated that, after correcting for sample selectivity in their decomposition analysis, the wage differential between White male and ethnic minority employees slightly increased from 10% to 11%.

As a sample selectivity check, I estimated the proportion of people in employment relative to those unemployed and the share of employees versus self-employed individuals (conditional on employment) by ethnic group and gender (see Tables 4.A5 and 4.A6 in the Appendix). Given that the sample comprises highly educated people, it is not surprising that the average employment rate is very high (96.8% for men and 97.5% for women). However, Black men are less likely to have a job (90.7%) than the rest of ethnic groups, particularly with respect to their White counterparts (97%). For women, the corresponding ethnic differences in the employment probability are smaller. On the other hand, the likelihood of being an employee (rather than self-employed) conditional on employment is identical across all male ethnic groups, standing at 84.5% on average (Table 4.A6). Black and Asian women are two percentage points more likely to be employees than their White counterparts. Taken together, the ethnic pay gaps estimated in the present analysis may, to a small extent, underestimate the effect of ethnicity on wages on the extensive margin if discrimination or other unobserved factors affect the economic activity and employment status.

Moreover, the APS datasets do not include variables capturing family characteristics, such as parental socio-economic background or income, and parental education[23]. These factors influence labour market outcomes and make up a significant source of difference between ethnic groups (Zuccotti, 2015). For example, parents from high social classes are more likely to transfer soft skills and help their children build social networks or support the young adults financially for more prolonged periods until the latter find a job that matches their preferences and educational profile (Zwysen and Longhi, 2018). Given that ethnic minority employees usually come from lower social backgrounds than White people (Platt, 2007; Zuccotti, 2015), not controlling for the parental socio-economic background would result in a downward bias in the effect of ethnicity on wages.

Additionally, some of the independent variables included in the OLS regression likely mirror responses to discrimination traced earlier in employees' career. This situation is described as "feedback effects" (Neumark, 2018) and reflects the interrelationship between wages and specific controlling characteristics (Gronau, 1988). For instance, the shorter firm tenure of women in the labour market may partly represent employers' presumptive expectations that

[23] *For the first time, the UK Labour Force Survey included a few social mobility questions in 2014 (such as parents' occupation when the respondent's age was 14 years). However, this information has not been incorporated in the APS datasets yet, possibly because those new questions are not consistently asked across all LFS waves (quarters).*

women are more likely to leave their job than men. As a result, employers do not equally invest in women acquiring firm-specific training, which would boost their productivity and, consequently, wages. Hence, this behaviour could, in turn, reduce women's motivation to remain in the company and impel them to quit their job, thus confirming employers' initial expectations. In the presence of feedback effects, conventional regression analysis might understate the impact of unexplained wage gaps associated with previous discrimination, thereby biasing the coefficients of ethnicity if such effects vary across ethnic groups.

I do not explicitly control for the above-mentioned factors because of data unavailability, which compromises the causal inference between ethnicity and wages. However, in section 4.5 (sensitivity analysis), I present the results from the partial identification method proposed by Oster (2017), which deals with the selection on unobservable characteristics. Finally, sample size restrictions for non-White workers prevent a detailed decomposition analysis by ethnic minority group for each gender or across different quantiles of the wage distribution, as this might result in unsound conclusions.

4.4 Results

Throughout this section, the coefficients on ethnic minority variables (Black, Asian, and Mixed/Other) are the headline figures[24]. Given that the "Mixed/Other" ethnic group is markedly heterogeneous, it is not the focal point in the following analysis.

4.4.1 Main OLS estimation

I start with a standard OLS estimation to examine the impact of being from an ethnic minority background on hourly wages, separately for men and women (Table 4.5). The results reveal substantial wage gaps between White employees and ethnic minorities, even after accounting for the entire pool of demographic, higher education and occupation/sector characteristics described in section 4.3 (see models 3 and 6 of Table 4.5). The earnings differential in favour of the White majority group is larger amongst men than women, but it decreases at a slower pace for men when successively including further controls in the regression. Specifically, the wage penalty for Black men ranges from -0.245 log wage points (-21.7%) in the baseline regression (model 1) to -0.183 log points (-16.7%) when

[24] The regression coefficients correspond to log wage differences, which are very close to real percentage differences for low values (<0.1). To transform the coefficients into real percentage effects, I use the formula $(e^{coefficient} - 1) * 100$.

adding in characteristics regarding higher education and occupation (model 3). UK-born male graduates from Asian minorities (that is, employees from Pakistani, Bangladeshi, Chinese, and any other Asian backgrounds) earn 4.1% less, on average, than their White counterparts, keeping all else constant. The wage penalty faced by the Mixed/Other ethnic group becomes statistically insignificant (-1.3%) when accounting for the complete set of observable determinants of earnings.

The story for women is contrasting. The earnings gap for Black women declines from -11.9% in the first regression to a three-time smaller effect of -4.5% in the full model. The ethnic penalties for Asian and Other/Mixed females stand at 2.0% and 3.5%, respectively. One explanation for the moderate ethnic penalties among women relative to men pertains to the disparate occupational and firm choices between genders, albeit differences remain even after controlling for occupation and industry groups. The corresponding figures in the Appendix (Tables 4.A3 and 4.A4) show that female graduates are more likely to work in the Public Admin/Education/Health sectors (58.1%), which are amongst the lowest-paid fields. This ties in with evidence from the literature suggesting that women have a greater probability of being employed in traditional “female occupations” (Manning and Swaffield, 2008; Lindley, 2016). In addition, a lower proportion of women (72.2%) enter higher-salaried jobs (“managerial/professional”) compared to men (80.4%). However, this is not explicit evidence of milder discrimination against ethnic minority females, as it could also signify the gender pay inequalities experienced by White women.

Another reason explaining the disparities in ethnic penalties between genders should be that women make different choices of the subject of study (Chevalier, 2011). For example, 27% of women in the present analysis sample have a first degree in the Arts/Humanities/Education areas, which yield lower labour market returns relative to other subjects. Moreover, the impact of unobserved characteristics on wages (including the magnitude of discrimination) may differ between men and women across occupations and subject areas of the first degree.

Analysing the effect of other independent variables on wages reveals some intriguing findings. As Table 4.5 shows, the “good degree” wage premium is 6.5% for men and 3.6% for women after controlling for all variables. This is comparable with the premium estimated by Naylor et al. (2016). Moreover, high-status university graduates enjoy, *ceteris paribus*, an average premium of approximately 10.5% across both genders. This resembles the Russell Group universities' premium found by Walker and Zhu (2018) before adjusting for the institution and course selectivity (measured by the average students' A-level scores within each university and subject of study). Interestingly, the results presented here show that the financial returns to attending a prestigious university exceed those to a good degree. This could reflect the pre-entry ability of students attending elite universities, the efforts that research-intensive institutions put on enhancing their students' productivity and career skills by offering, for example, occupational opportunities with partner firms (Russell Group, 2019), or it could also be a corollary of signalling effects of having graduated from an elite institution (Spence, 1973; Belfield et al., 2018a).

Male employees holding a first degree in Engineering/Technology subjects attract an earnings advantage of 5.9% compared to the Social studies graduates, whereas the corresponding pay disadvantage for those graduating from Arts/Humanities/Education stands at 10.7% relative to the omitted category. Except for the subjects allied to Law/Business/Finance, which yield a wage reward of 5.7% compared to Social studies, the differences in subject returns are smaller amongst women than men.

The managerial/professional occupations offer an impressive average wage advantage of 37.9% for men and 41.8% for women, keeping all else equal. In a similar vein, medium/large companies' personnel enjoy notably higher salaries than the employees of small/micro enterprises. Furthermore, working in the Manufacturing/Construction and Financial sectors improves earnings for both genders. Part-time jobs bring an average wage penalty of 8.2% for men (relative to full-time employment), but no statistically significant differences are established for women. Across both genders, permanent jobs are linked with higher earnings prospects than fixed-time contracts and seasonal work. Male employees working in the public sector earn 6.9% less than those working for private firms, whereas women receive a small pay benefit of 1.2% if they work for public organisations. The workplace's geographical location also plays a crucial role, with London wages being significantly higher relative to the rest of UK regions. Unsurprisingly, years of firm tenure are positively correlated with wages, and the effect of this variable is more prominent amongst women, while working more hours per week also imparts higher earnings.

Being married (or in a civil partnership) comes with positive consequences on wages, both for men (8.9%) and women (5.8%). Although married people exhibit better earnings levels, this does not necessarily imply that marriage *per se* raises wages. Instead, partnered employees may have other unobserved characteristics that are correlated with increased earnings (e.g., economic motivation). Finally, having children is associated with a 5.4% rise in wages for men. In contrast, consistent with other studies evincing discrepancies in earnings between females with dependent children and those without children (e.g., Gangl and Ziefle, 2009; Viitanen, 2012), the “motherhood wage penalty” stands at 1.1%, keeping all other wage determinants fixed. Waldfogel (1998) linked this “family gap” to unmeasured heterogeneity (e.g., mothers may have lower incentives for market work and career advancement than women without children), discrimination from employers against mothers, and institutional hurdles surrounding the participation of mothers in the workforce (such as rigid working times and deficient childcare policies).

In sum, the regression results document strong ethnic pay gaps in favour of White employees, particularly amongst men, even after accounting for a variety of characteristics that influence wages. The earnings differential is enormous for the Black community, who are strikingly penalised in the UK labour market relative to equally qualified White workers.

Table 4.5. OLS regressions: by gender

Dependent variable: log(wage)

	Men			Women		
Variable	(1) Demographics	(2) +Subjects, University	(3) +Occupation, Sector	(4) Demographics	(5) +Subjects, University	(6) +Occupation, Sector
Black	-0.245*** (0.032)	-0.217*** (0.031)	-0.183*** (0.029)	-0.127*** (0.021)	-0.089*** (0.021)	-0.046** (0.018)
Asian	-0.044*** (0.016)	-0.060*** (0.016)	-0.042*** (0.014)	-0.036** (0.014)	-0.035** (0.014)	-0.020 (0.012)
Mixed/Other	-0.052* (0.028)	-0.050* (0.026)	-0.013 (0.023)	-0.069*** (0.021)	-0.067*** (0.021)	-0.036* (0.018)
Age	0.095*** (0.002)	0.096*** (0.002)	0.071*** (0.002)	0.092*** (0.002)	0.092*** (0.002)	0.062*** (0.002)
Age squared	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
London	0.342*** (0.010)	0.332*** (0.010)	0.267*** (0.009)	0.321*** (0.009)	0.304*** (0.009)	0.244*** (0.009)
South East	0.111*** (0.009)	0.107*** (0.009)	0.087*** (0.009)	0.071*** (0.008)	0.066*** (0.008)	0.060*** (0.007)
Northern	-0.028*** (0.009)	-0.026*** (0.008)	-0.020** (0.008)	0.001 (0.007)	0.000 (0.007)	-0.001 (0.006)
Wales	-0.109*** (0.012)	-0.107*** (0.011)	-0.077*** (0.010)	-0.021** (0.009)	-0.038*** (0.009)	-0.036*** (0.008)
Scotland	0.007 (0.010)	-0.012 (0.010)	0.004 (0.009)	0.070*** (0.008)	0.052*** (0.008)	0.041*** (0.007)
N. Ireland	-0.098*** (0.021)	-0.123*** (0.021)	-0.093*** (0.020)	-0.036** (0.016)	-0.062*** (0.017)	-0.065*** (0.015)
Partnered	0.132*** (0.007)	0.127*** (0.007)	0.085*** (0.007)	0.088*** (0.006)	0.081*** (0.006)	0.056*** (0.005)
Any dependent children in family aged <19	0.054*** (0.007)	0.050*** (0.007)	0.053*** (0.007)	-0.060*** (0.006)	-0.057*** (0.006)	-0.011* (0.005)
Health problems lasting >1 year	-0.065*** (0.008)	-0.053*** (0.007)	-0.032*** (0.007)	-0.051*** (0.006)	-0.044*** (0.006)	-0.025*** (0.005)
Health		0.061*** (0.018)	0.054*** (0.017)		0.117*** (0.009)	0.021** (0.009)
Sciences		0.037*** (0.012)	-0.010 (0.011)		0.013 (0.010)	-0.016* (0.009)
Engineering/ Technology		0.139*** (0.013)	0.057*** (0.013)		0.099*** (0.019)	0.018 (0.017)
Law/Business/Finance		0.071*** (0.013)	0.032*** (0.012)		0.079*** (0.010)	0.055*** (0.009)
Arts/Humanities/ Education		-0.137*** (0.013)	-0.113*** (0.012)		-0.028*** (0.009)	-0.023*** (0.008)
Combined subject		0.020 (0.014)	-0.004 (0.012)		0.023** (0.010)	0.011 (0.009)
High-status universities		0.130*** (0.006)	0.100*** (0.006)		0.130*** (0.006)	0.098*** (0.005)
Good degree		0.091*** (0.006)	0.063*** (0.006)		0.063*** (0.005)	0.035*** (0.005)
Managerial/ Professional job			0.321*** (0.007)			0.349*** (0.006)
Public Admin/Education/ Health			-0.011 (0.013)			-0.012 (0.011)

Continued on next page

Table 4.5. (continued)						
Banking/Finance			0.092*** (0.012)			0.101*** (0.012)
Trade/Hotel/Restaurant			-0.047*** (0.014)			-0.062*** (0.013)
Transport/ Communication			0.101*** (0.013)			0.095*** (0.015)
Manufacturing/ Construction			0.081*** (0.013)			0.127*** (0.014)
Medium/Large enterprises (>=50 employees)			0.153*** (0.006)			0.078*** (0.005)
Public sector			-0.072*** (0.009)			0.012* (0.006)
Part-time work			-0.086*** (0.015)			-0.011 (0.008)
Permanent job			0.058*** (0.016)			0.022** (0.010)
Years in current employer (Tenure)			0.004*** (0.000)			0.009*** (0.000)
Total usual weekly hours in main job			0.002*** (0.000)			0.003*** (0.000)
Constant	0.620*** (0.043)	0.508*** (0.045)	0.599*** (0.047)	0.706*** (0.036)	0.602*** (0.038)	0.771*** (0.038)
Survey year/Survey month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,288	21,713	21,370	27,078	26,321	25,900
Adjusted R^2	0.294	0.339	0.448	0.195	0.228	0.417

Note: Robust standard errors in parentheses.
** p < 0.1, ** p < 0.05, *** p < 0.01.*
The regression sample comprises UK-born employees (aged 19-65) who hold a first degree from a UK university. It excludes individuals with a higher degree (e.g., Masters, Doctorate).
The independent variables included in the regression specification are: Ethnicity (Black, Asian, Mixed/Other), Age, Age squared, Region of workplace (London, South East, Northern, Wales, Scotland, N. Ireland), Partnered, Any dependent children in family aged <19, Health problems lasting >1 year, Subject area of first degree (Health, Sciences, Engineering/Technology, Law/Business/Finance, Arts/Humanities/Education, Combined subject), High-status universities, Good degree, Managerial/Professional job, Industry sector (Public Admin/Education/Health, Banking/Finance, Trade/Hotel/Restaurants, Transport/Communication, Manufacturing/Construction), Medium/Large enterprises (>=50 employees), Public sector, Part-time work, Permanent job, Years in current employer (Tenure), Total usual weekly hours in main job, Survey year dummies, Survey month dummies.
The base categories for the multi-categorical dummy variables are: "White" (for ethnicity), "Rest of England" (for the region of workplace), "Social studies" (for the subject area of first degree) and "Other sector" (for the industry sector).
Source: APS 2013-2018

4.4.2 Heterogeneous effects

4.4.2.1 Quantile regression estimation

This subsection examines the heterogeneous effects of ethnicity on wages calculated at different levels of the dependent variable (log hourly wage) based on the quantile regression estimates. The regressions presented here include the same explanatory variables as in the previous section (4.4.1).

The following results should be interpreted cautiously because the conditional quantile regression estimates cannot be generalised to the entire population distribution (Firpo, Fortin and Lemieux, 2009). Specifically, the estimated effect of ethnicity on each quantile of wages is conditional on specific values of the other independent variables (such as age, region, “good” degree, subject of study, and university type). In other words, for each level of the other covariates, this approach generates a conditional wage distribution, and each quantile may represent a mix of different points of the unconditional distribution. Hence, the estimated effect of ethnicity on earnings at a conditional quantile is a weighted average of the corresponding effects within each subgroup defined by the other regressors (that is, *age*region*good degree*subject of study*, and so forth). However, the groups that contribute to the average estimated effects at each conditional quantile may lie in different segments of the unconditional earnings distribution. Therefore, the ethnic penalties reported in the present subsection reflect within-group inequalities and cannot be extrapolated to the overall earnings distribution (that is, the unconditional distribution).

Figures 4.8 and 4.9 depict the wage gaps for the ethnic minority groups (relative to their White counterparts) across nine conditional quantiles (segments) of the log wage distribution for men and women, respectively. Similarly, Tables 4.A7 and 4.A8 in the Appendix report the detailed quantile regression estimates at the 0.10, the median and the 0.90 cut-off points of the conditional earnings distribution for each gender separately. The results confirm that there is heterogeneity in wage gaps between White and ethnic minority employees at different earnings levels. Specifically, the ethnic penalties in wages for Black men remain relatively stable between the 10^{th} and 80^{th} quantiles of the conditional earnings distribution, ranging around 16%-18% (from -0.17 to -0.20 log points). However, the magnitude of the wage differential between White and Black men drops to 11.7% when looking at the upper (90^{th}) quantile. On the contrary, the pay inequalities seem to worsen with earnings for Black women, as they experience a sharply increased wage gap of 9.8% at the top quantile of the conditional distribution compared to White females (relative to 4.0% and 3.4% at the 0.10 and the median points, respectively).

The acceleration of the pay gap in the upper tail of the conditional earnings distribution for Black women might indicate the presence of the “glass ceiling” phenomenon[25] (Albrecht, Björklund and Vroman, 2003). This could be the consequence of various reasons, such as discrimination in promotion procedures (regarding the advancement probability or monetary rewards to promotion),

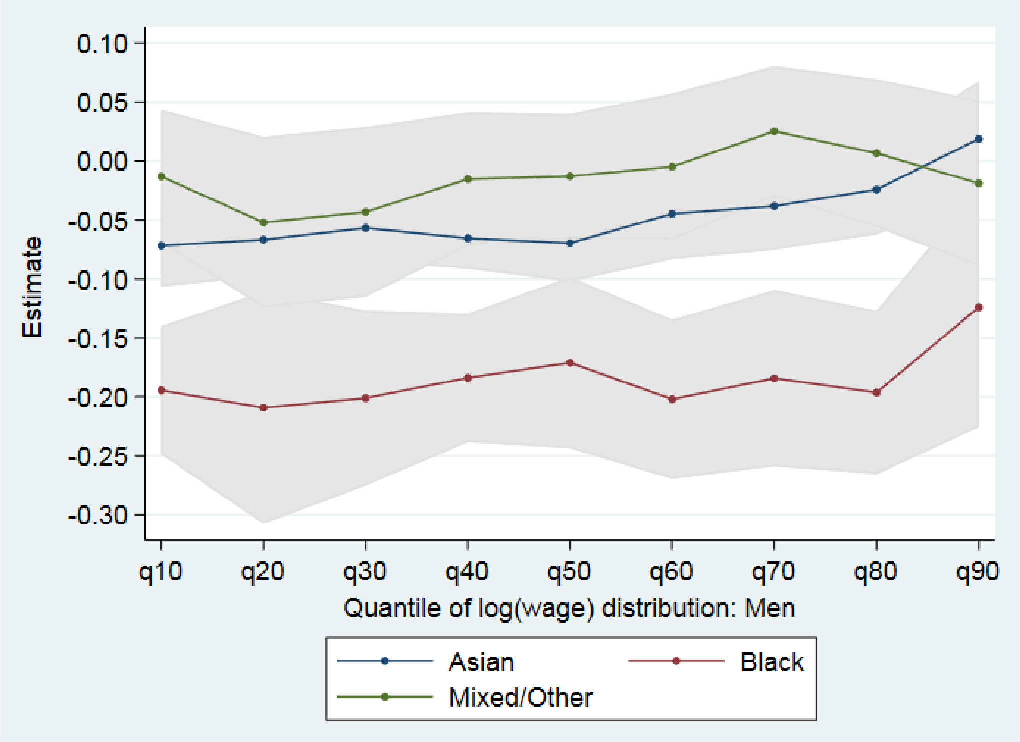
[25] *However, as mentioned earlier, the higher part of the conditional wage distribution does not necessarily mirror the higher part of the unconditional distribution.*

decreased chances of working in companies that offer high pay premiums and on-the-job training (which would improve Black women's human capital after being recruited), or diverse preferences to compete (Niederle and Vesterlund, 2007; Vlassopoulos and Siddique, 2020), which might stop them from pursuing positions of authority. Moreover, a higher proportion of Black women might work in less-demanding jobs (if, for example, they have greater childcare commitments) or demonstrate weaker salary negotiation skills than White females (see a discussion about how these factors influence the gender gaps in Arulampalam, Booth and Bryan, 2007; Card, Cardoso and Kline, 2016). I cannot investigate these potential mechanisms using the data at hand. Nonetheless, they should, at least partially, explain the previous results.

The pay gaps for Asian men substantially improve when moving up to higher segments of the conditional earnings distribution (from approximately -6.9% at the median and the bottom quantiles to a positive but statistically insignificant effect of 1.9% at the 90th quantile). The wider ethnic pay differences for Asian men at the lower paying parts of the conditional wage distribution might provide supportive evidence of the "sticky floor" hypothesis (Booth, Francesconi and Frank, 2003), although the "floor" could be somewhere different in the unconditional wage distribution for a given set of other covariates. This plausibly suggests that employers offer better starting salaries to White men than their Asian counterparts, particularly in low-paying firms and sectors, even if both groups are recruited at equivalent pay bands (but Asian people are sorted more into the bottom of the relevant scales). However, as Asian males make their way up the career ladder, they catch up to the White workers' earnings rates. In contrast, the wage differences between White and Asian women are statistically insignificant across most quantiles of the conditional wage distribution, peaking at 4.4% at the 80th quantile.

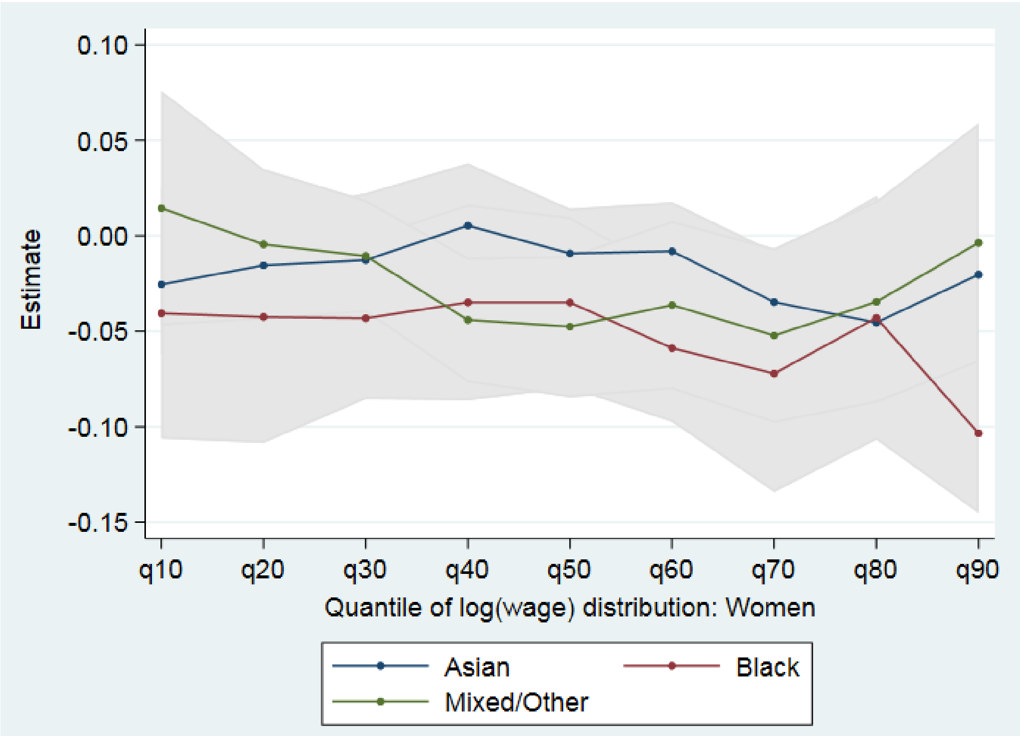
For men, there are no discernible patterns in earnings differentials for the remarkably heterogeneous Mixed/Other ethnic group, as wage gaps remain small and statistically insignificant throughout the conditional earnings distribution. Conversely, the pay gaps for women from Mixed/Other backgrounds are more pronounced around the median (-4.7%) but vanish at the top and bottom parts of the wage distribution.

Figure 4.8. Ethnic wage gaps across quantiles of the wage distribution – Men



Note: The estimates shown in the graph are derived from the corresponding quantile regressions. The shaded areas represent the 95% confidence intervals (CI) of the estimated coefficients. CI are calculated based on bootstrapped standard errors (100 repetitions), which are robust to heteroskedasticity.
Source: APS 2013-2018

Figure 4.9. Ethnic wage gaps across quantiles of the wage distribution – Women



Note: The estimates shown in the graph are derived from the corresponding quantile regressions. The shaded areas represent the 95% confidence intervals (CI) of the estimated coefficients. CI are calculated based on bootstrapped standard errors (100 repetitions), which are robust to heteroskedasticity.
Source: APS 2013-2018

4.4.2.2 OLS regressions within subgroups of employees

Table 4.6 presents the coefficients of ethnicity derived from separate OLS regressions within sub-samples of employees. Specifically, I explore whether the ethnic penalties vary by the subject area of study (STEM, LEM, Other, Combined subjects), the type of university (high status, other institutions), the degree class (good degree, other class of degree), the occupation group (managerial/professional, other jobs), the workplace size (micro/small enterprises, medium/large firms), and age bands (19-30 years, 31-65 years). Although I do not show the complete list of coefficients in this table to save space, the regressions include all controls encompassed in Table 4.5.

For men, there are economically significant differences in the magnitude of ethnic inequalities in earnings across the main subject areas of the first degree and between institution types. Holding all other factors fixed, Black male graduates from STEM and Combined subjects are penalised by 19.3% and 21.2%, respectively, compared to their White counterparts (Panel A). The wage gaps for Black men are considerably smaller within the "other" subject areas (-5.8%), which include, among others, degrees related to Languages, Humanities, Arts, Education and Mass Communication. Similarly, Black men who graduate from a lower-status university (Panel C) experience a 17.4% wage gap, which is three percentage points higher than the wage difference observed among alumni of high-status institutions. Black men with a good degree see higher earnings penalties than those attaining a lower class of degree (Panel D).

Interestingly, for Black men, the wage differential is more pronounced within the well-compensated occupations (managerial/professional) than other jobs, as the difference in pay gaps between these two broad occupational groups is 6.5 percentage points (Panel E). Likewise, Black male employees experience substantially higher pay disadvantages within the medium- or large-sized firms (Panel F). Collectively, these figures likely corroborate a picture of more intense discrimination against Black men in well-paid occupations and high-paying companies.

Conversely, in line with the findings from the quantile regressions discussed earlier, the ethnic penalties for Asian men are more prominent among employees with lower expected wages. Specifically, the wage gaps for Asian males are nearly two times higher for those graduating from a non-high-status university, holding a lower degree class, having a non-managerial/professional job, and being employed in small/micro enterprises. These findings could imply that White men systematically end up in companies that offer higher salaries and pay premiums relative to Asian men, even if both groups work within the same industries and hold similar positions. These results might also suggest that employers value some higher education characteristics of Asian men (such as the degree class, which acts as a productivity signal at early career stages) more favourably than those of Black employees. The underlying mechanisms that drive these disparities for Asian men require future research, particularly using more extensive datasets to distinguish between the main ethnic subgroups (Chinese, Indian, Pakistani, and Bangladeshi) of the diverse Asian category.

Black women endure more substantial ethnic penalties if they hold a combined degree (-10.3%, compared to an average difference of -4.5% in the total sample), graduate from a prestigious university (-7.4%), or if they work for a medium/large-sized company (-4.9%). On one of the few occasions where earnings gaps favour ethnic minority employees, Asian women holding a STEM degree reap a wage premium of 4.9% compared to their White counterparts. In contrast, Asian female employees enrolled in a LEM course suffer a 7.0% pay disadvantage in the labour market. Employees from Mixed/Other ethnic backgrounds experience higher pay gaps amongst the good-degree holders (for women) and those who work for micro/small businesses (for men).

Finally, ethnic penalties seem to exacerbate with age for men and women (Panel G). In particular, there are no wage differences for women aged 30 and under across all ethnic groups. In contrast, remarkable ethnic pay gaps exist for older female employees (aged 31-65), ranging from 3.5% for Asian women to 6.1% for Black women (relative to their White counterparts), other things equal. Ethnic penalties persist for Black men across both age groups, but the wage gap size doubles from 10.1% in early career stages to 19.4% for those over 30 years old. The average wages conditional on the total pool of observed characteristics remain equal among young Asian and White males. However, older White men earn 6.1% more than Asian minorities, on average.

The smaller racial inequalities in the early phases of the graduates' life documented in this paper are generally consistent with the findings of Zwysen and Longhi (2018) in the UK. Hence, it appears that the ethnic wage disparities established in the literature develop through the career, and, as the present study reveals, they hold even among highly educated individuals. Thomas, Herring and

Horton (1994) used data for different cohorts during 1940-1990 in the US to show that the male Black-White pay gap is much smaller among young adults than middle-aged employees, and it converges among the elderly workers. The rapid increase of ethnic pay disparities over the first twenty years of Black men's employment remained for all cohorts throughout their analysis. Therefore, the authors disputed the "legacy of past discrimination" theory, which implies that any pay inequalities among ethnic groups exclusively echo past racial bias experienced by elder minorities, and these disparities should disappear for younger cohorts.

There may be several factors at play explaining why differences in wages between ethnic groups grow over the career. Tomaskovic-Devey, Thomas and Johnson (2005) asserted that some human capital characteristics (which affect the earnings level) are endogenous to the job market due to the interaction and social negotiation between firms and employees. In particular, specific human capital assets (such as experience and firm-specialised skills) are obtained within the labour market, and they do not merely depend on investment choices made by employees and their family. Hence, if there is discrimination against ethnic minorities relative to similarly qualified White people in terms of recruitment, promotion procedures, company tenure, access to training, and time needed to find a job (which affects the cumulative experience), there would be negative consequences for the ethnic wage gaps in the long term. For instance, promotions are positively associated with pay rises, training prospects, authority, and work satisfaction (Yap and Konrad, 2009).

Using a series of experiments and online-simulated negotiations in the US, Hernandez et al. (2019) showed that employers anticipate that Black candidates negotiate less than comparable White people. However, the authors demonstrated that Black applicants experience more severe penalties regarding the offered salaries when they violate the employers' initial stereotypical perceptions, and these penalties worsen when bargaining with more ethnically biased evaluators, who believe that Black people merit lower wages. Hence, from this perspective, if some ethnic minorities are systematically less willing to negotiate their wages due to perceived discrimination, then this should also partly explain their pay disadvantage relative to White employees.

Table 4.6. Heterogeneous effects: Separate OLS regressions by gender
Dependent variable: log(wage)

Panel A. OLS regression by subject area of study: **Men**

	All	STEM	LEM	Other	Combined
Black	-0.183***	-0.215***	-0.188***	-0.060	-0.238***
	(0.029)	(0.047)	(0.052)	(0.068)	(0.076)
Asian	-0.042***	-0.042*	-0.047**	0.001	-0.043
	(0.014)	(0.022)	(0.022)	(0.051)	(0.044)
Mixed/Other	-0.013	0.025	-0.017	-0.048	-0.084
	(0.023)	(0.035)	(0.043)	(0.060)	(0.059)
Observations	21,370	9,814	5,003	3,710	2,843
Adjusted R^2	0.448	0.415	0.436	0.454	0.454

Panel B. OLS regression by subject area of study: **Women**

	All	STEM	LEM	Other	Combined
Black	-0.046**	-0.041	-0.037	-0.041	-0.109***
	(0.018)	(0.040)	(0.032)	(0.036)	(0.037)
Asian	-0.020	0.048**	-0.073***	-0.032	-0.024
	(0.012)	(0.024)	(0.019)	(0.027)	(0.031)
Mixed/Other	-0.036*	-0.045	-0.046	-0.028	-0.025
	(0.018)	(0.032)	(0.037)	(0.034)	(0.052)
Observations	25,900	8,697	6,330	6,995	3,878
Adjusted R^2	0.417	0.385	0.444	0.438	0.417

Panel C. OLS regression by type of university

	Men			**Women**		
	All	High-status universities	Other universities	All	High-status universities	Other universities
Black	-0.183***	-0.154**	-0.191***	-0.046**	-0.077	-0.030
	(0.029)	(0.073)	(0.031)	(0.018)	(0.048)	(0.020)
Asian	-0.042***	-0.023	-0.054***	-0.020	-0.006	-0.020
	(0.014)	(0.027)	(0.017)	(0.012)	(0.024)	(0.014)
Mixed/Other	-0.013	-0.006	-0.017	-0.036*	-0.057	-0.023
	(0.023)	(0.041)	(0.028)	(0.018)	(0.039)	(0.020)
Observations	21,370	7,519	13,851	25,900	8,124	17,776
Adjusted R^2	0.448	0.403	0.458	0.417	0.384	0.419

Panel D. OLS regression by class of first degree

	Men			**Women**		
	All	Good degree	Other degree class	All	Good degree	Other degree class
Black	-0.183***	-0.202***	-0.164***	-0.046**	-0.048**	-0.041
	(0.029)	(0.041)	(0.040)	(0.018)	(0.024)	(0.028)
Asian	-0.042***	-0.033*	-0.060***	-0.020	-0.005	-0.037*
	(0.014)	(0.019)	(0.022)	(0.012)	(0.016)	(0.019)
Mixed/Other	-0.013	-0.008	-0.024	-0.036*	-0.061**	0.005
	(0.023)	(0.029)	(0.037)	(0.018)	(0.026)	(0.024)
Observations	21,370	11,452	9,918	25,900	15,571	10,329
Adjusted R^2	0.448	0.464	0.432	0.417	0.419	0.421

Continued on next page

Table 4.6. (continued)						
Panel E. OLS regression by major occupation group						
	Men			**Women**		
	All	Managerial/ Professional	Other occupation	All	Managerial/ Professional	Other occupation
Black	-0.183***	-0.201***	-0.125**	-0.046**	-0.051**	-0.059*
	(0.029)	(0.033)	(0.062)	(0.018)	(0.023)	(0.030)
Asian	-0.042***	-0.035**	-0.063**	-0.020	-0.014	-0.030
	(0.014)	(0.016)	(0.028)	(0.012)	(0.015)	(0.020)
Mixed/Other	-0.013	-0.023	0.034	-0.036*	-0.037	-0.039
	(0.023)	(0.026)	(0.047)	(0.018)	(0.024)	(0.025)
Observations	21,370	17,222	4,148	25,900	18,751	7,149
Adjusted R^2	0.448	0.339	0.406	0.417	0.256	0.293
Panel F. OLS regression by workplace size						
	Men			**Women**		
	All	Micro/Small enterprises	Medium/ Large	All	Micro/Small enterprises	Medium/ Large
Black	-0.183***	-0.122**	-0.205***	-0.046**	-0.034	-0.050**
	(0.029)	(0.055)	(0.033)	(0.018)	(0.029)	(0.023)
Asian	-0.042***	-0.056**	-0.029*	-0.020	-0.022	-0.024*
	(0.014)	(0.028)	(0.016)	(0.012)	(0.023)	(0.014)
Mixed/Other	-0.013	-0.071**	0.027	-0.036*	-0.030	-0.040*
	(0.023)	(0.035)	(0.030)	(0.018)	(0.030)	(0.023)
Observations	21,370	7,647	13,723	25,900	10,568	15,332
Adjusted R^2	0.448	0.371	0.471	0.417	0.398	0.415
Panel G. OLS regression by age bands						
	Men			**Women**		
	All	19-30 years	31-65 years	All	19-30 years	31-65 years
Black	-0.183***	-0.106**	-0.216***	-0.046**	0.004	-0.063***
	(0.029)	(0.046)	(0.036)	(0.018)	(0.030)	(0.023)
Asian	-0.042***	0.005	-0.063***	-0.020	0.006	-0.036**
	(0.014)	(0.021)	(0.019)	(0.012)	(0.017)	(0.017)
Mixed/Other	-0.013	0.047	-0.045	-0.036*	0.001	-0.059**
	(0.023)	(0.032)	(0.031)	(0.018)	(0.023)	(0.027)
Observations	21,370	5,683	15,687	25,900	7,734	18,166
Adjusted R^2	0.448	0.450	0.322	0.417	0.437	0.337

Note: Robust standard errors in parentheses.
** p < 0.1, ** p < 0.05, *** p < 0.01.*
Although only the coefficients of ethnic groups are reported in this table, all regressions also comprise all controls presented in Table 4.5. Specifically, the independent variables included in the regression specification are: Ethnicity, Age, Age squared, Region of workplace, Partnered, Any dependent children in family aged <19, Health problems lasting >1 year, Subject area of first degree, High-status universities, Good degree, Managerial/Professional job, Industry sector, Medium/Large enterprises (>=50 employees), Public sector, Part-time work, Permanent job, Years in current employer (Tenure), Total usual weekly hours in main job, Survey year dummies, Survey month dummies. The regression sample comprises UK-born employees (aged 19-65) who hold a first degree from a UK university. It excludes individuals with a higher degree (e.g., Masters, Doctorate).
Source: APS 2013-2018

4.4.3 Decomposition analysis

This subsection explores the drivers of pay gaps between White and non-White employees by quantifying the contribution of each group of individual characteristics to the earnings differential. Table 4.7 displays results of the decomposition analysis at the mean of the earnings distribution. The average wage difference between White and non-White employees is 0.061 log points. The results show that the observed characteristics (explained component) drive nearly 30% (0.018 log points) of the total wage gap. However, the largest part of the wage differential (0.043 log points or 70%) is attributable to ethnic discrimination or other unobservable factors correlated with ethnicity and wages.

Further decomposing the explained component into “demographics”, “higher education” and “occupation/sector” characteristics results in some interesting findings. Specifically, discrepancies in the average observed occupation/sector-related factors between White and ethnic minority workers (0.031 log points) account for 51% of the total earnings differential. Put differently, if the inequalities related to job characteristics (such as the uneven representation of non-White employees in managerial/professional occupations and well-paying sectors) are addressed, the wage gap could shrink by up to 51% to the benefit of ethnic minorities (on average for both genders).

Figure 4.10 depicts the proportional contribution of each group of observed characteristics (on average) to the wage differential, as a whole and separately for each gender. The first column (“wage gap”) represents the raw difference in hourly earnings between White and non-White employees, while the second bar (“explained part”) illustrates by how much the wage difference would decrease if ethnic minority employees had similar observed characteristics to the White workers. The following bars of the graph decompose the explained component of the earnings differential into three parts, as described above. The positive figures in the graph mean that the corresponding factors contribute to the ethnic wage penalties, whereas the negative ones work in the opposite direction and mitigate the pay gaps. Specifically, the negative values indicate that some characteristics of ethnic minorities reduce the ethnic wage gap, implying that the earnings differential would be even greater if both ethnic groups (White and non-White people) had the same characteristics.

For instance, as discussed earlier, many ethnic minority employees are concentrated in London, where the average wages are remarkably higher than in the other regions of the UK. The detailed results[26] reveal that if the share of White people working in London was identical to that of ethnic minorities, then the wage gap would be nearly six percentage points higher for non-White employees. However, the younger age profile of ethnic minority employees is the key factor that counterbalances the “London effect”, thus resulting in a net effect of -1.4% of the demographics factors in the total explained wage differential. On the contrary, the fewer years of job tenure and the unequal participation in higher-paid occupations (managerial/professional jobs) are the main contributing determinants of the ethnic wage gap associated with the occupation/sector group of characteristics (3.1%). Dissimilarities in higher education characteristics matter little on average for both genders.

Delving deeper into the results by focusing on each gender separately, the conclusions are somewhat different. For men, the explained part (0.030 log points) accounts for 34% of the total wage gap, while the corresponding share for women stands at 17% (0.006 log points). Job-pertinent characteristics continue to cover most of the explained component for men (87%). The remaining explained differences in wages arise from demographic factors, such as the younger age profile of ethnic minorities (compared to White male employees) and marriage patterns, which offset the positive effect of working in London. In contrast, the demographic traits have an opposite effect on women’s wages, as, on average, they reduce pay gaps in favour of non-White females. Specifically, although non-White women are younger (relative to White females), the effect of age is lower than that of men and is outweighed by other demographic factors (such as the region of the workplace, which, as mentioned earlier, usually tallies with the place of residence). Nevertheless, the job/sector-related characteristics (0.036 log points) remain the most critical determinants of the explained part of ethnic pay gaps amongst women.

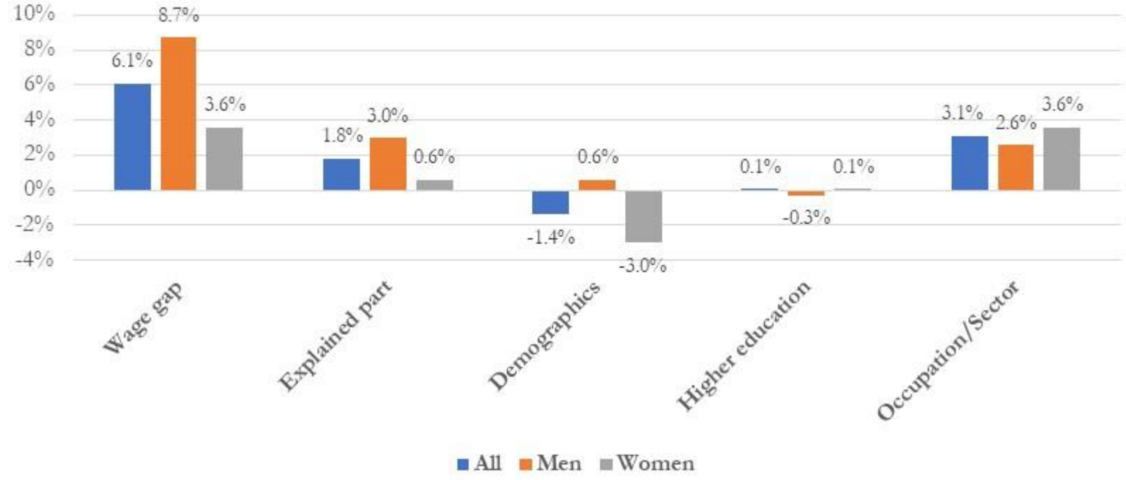
In a nutshell, the occupational segregation, expressed by the under-representation of ethnic minorities in high-salaried occupations and well-paying sectors, in combination with their shorter job tenure relative to White people, constitute the primary drivers of pay gaps relating to observed characteristics. Despite this, the unobserved characteristics and possible discriminatory practices

---

[26] *Due to space constraints and for illustration purposes, in the present decomposition analysis, I do not report the individual contribution of the variables that form each group of characteristics.*

in employees' compensation and promotion represent the grand part of the wage differential across both genders.

Figure 4.10. Decomposition analysis: impact of observed characteristics on ethnic wage gaps (White vs. non-White employees)



Note: The graph figures are derived from Table 4.7. Given that most coefficients of Table 4.7 are small, they are very close to real percentage effects and I, therefore, did not apply the correction presented in section 4.4.1.
Source: APS 2013-2018.

Table 4.7. Decomposition analysis

Log (wage)	(1) All	(2) Men	(3) Women
White	2.814	2.933	2.715
Non-White	2.753	2.846	2.679
Difference	0.061***	0.087***	0.036**
	(0.010)	(0.015)	(0.012)
Explained component			
Demographics	-0.014**	0.006	-0.030***
	(0.004)	(0.008)	(0.005)
Higher education	0.001	-0.003	0.001
	(0.001)	(0.002)	(0.002)
Occupation/Sector	0.031***	0.026***	0.036***
	(0.004)	(0.006)	(0.006)
Total	0.018*	0.030**	0.006
	(0.007)	(0.011)	(0.009)
Unexplained component			
Demographics	-0.210	-0.125	-0.133
	(0.124)	(0.206)	(0.154)
Higher education	0.004	0.015	0.003
	(0.005)	(0.008)	(0.009)
Occupation/Sector	-0.006	-0.038	0.002
	(0.041)	(0.073)	(0.049)
Constant	0.255*	0.205	0.158
	(0.129)	(0.211)	(0.159)
Total	0.043***	0.057***	0.029**
	(0.007)	(0.012)	(0.009)
Observations	47,270	21,370	25,900

Note: Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Each group includes the following variables:
"Demographics": Region of workplace, Age, Age squared, Partnered, Any dependent children in family aged <19, Health problems lasting >1 year.
"Higher education": Subject area of study (7 categories), High-status university, Good degree.
"Occupation/Sector": Managerial/Professional job, Industry sector (6 categories), Workplace size, Public sector, Part-time work, Permanent job, Tenure, Total usual weekly hours in main job.
Source: APS 2013-2018

4.5 Sensitivity analysis

This section explores the sensitivity of the effect ethnicity has on wages based on: i) various matching techniques; ii) the coefficient stability method proposed by Oster (2017), and iii) further robustness checks.

4.5.1 Matching methods

In the cases where there is a treated and a control group, matching techniques aim to overcome the selection bias on observable factors by balancing the distribution of the covariates between both groups (Caliendo and Kopeinig, 2005). In the present setup, the treatment group consists of employees from a specific ethnic minority (e.g., Black people), and the control group comprises their White counterparts. The fundamental idea of the semi-parametric matching methods is to detect in a broad group of White workers those who are identical to the specified ethnic minority group (e.g., Black employees) in terms of observed characteristics. In doing so, these methods adjust for differences in the mean and variance (or higher moments) of the independent variables' distribution, thus producing a sample of "statistical twins" between both groups. For example, one would like to predict the average wages of Black employees if they had exactly the same traits as their White counterparts but only differed in their ethnicity. Because it is impossible to observe both the actual and counterfactual wage outcomes for the same person simultaneously, the matching methods are a common solution to the selection issue in empirical applications (Crawford et al., 2010).

The traditional parametric approaches, such as the OLS, are based on assumptions about the functional form of the regression specification. However, the precise functional form of the association between the dependent variable (log hourly wage) and the covariates is usually unknown. For example, omitting the quadratic term of an explanatory variable, which determines the outcome in the true population model, would bias the regression estimates. Matching techniques reduce the impact of the researcher's modelling choices ("model dependence"), which are often not justified from the economic theory (Jann, 2017a). Rosenbaum and Rubin's (1983) pioneering paper developed the matching algorithms on the basis of a propensity score function. This parametric function estimates the likelihood of being treated (that is, being from a specific ethnic minority background) conditional on the observed characteristics. Hence, the information about all independent (observable) variables is summarised into a single function (propensity score), instead of calculating a practically infeasible highly dimensional vector of distinct covariates for each sample unit. Because this function is not directly correlated with the outcome (that is, wages), it is more robust to misspecification issues (Huber, Lechner and Wunsch, 2013).

Another disadvantage of the OLS method is that it calculates the missing counterfactual by extrapolating to dissimilar individuals when the common support requirement fails. Specifically, OLS may project the earnings distribution of White employees onto areas of the ethnic minorities' distribution where the proportion of the latter is negligible or virtually nil (Barsky et al., 2002; Jurajda and Paligorova, 2009). Matching methods overcome this shortcoming by dropping unmatched sample units or applying weight functions to observations to obtain a sample of comparable treatment and control groups.

The average treatment effect on the treated (ATT) denotes the expected difference between the actual and the potential (counterfactual) wage outcomes for the subpopulation of employees that compose the treated group (that is, for each ethnic minority in the current setting). More formally,

$$ATT = E[Y_{i1} - Y_{i0} | T = 1] = E[Y_{i1} | T = 1] - E[Y_{i0} | T = 1] \quad (4)$$

where Y_{i1} and Y_{i0} are the potential conditional wage outcomes of the ethnic minority employee i with treatment $(T = 1)$ and without treatment $(T = 0)$, respectively. Clearly, $E[Y_{i0} | T = 1]$ cannot be observed in the data as it describes the counterfactual outcome (that is, the hourly wage of employee i if they were White). If all earnings determinants are observed in the data, then the ATT represents the causal effect of each ethnic group on wages.

However, matching techniques impose a strong requirement to establish causality — the "conditional independence assumption" (CIA). This assumption states that the potential values of the dependent variable (log hourly wages) if an individual is treated and the corresponding values if that individual is not treated are independent of the treatment, conditional on the covariates. In other words, taking the example of comparing the wages between Black and White employees, the CIA requires that the wages of White workers should be equal to the unobserved wages of Black people with the same characteristics if the latter were White, thus assuming a random allocation of the treatment. As a corollary, the likely presence of unobserved differences between the ethnic groups, which also influence wages, would bias the matching estimates, even when a maximum post-match balance between the treatment and control groups is achieved (common support assumption).

Despite these limitations, non-experimental studies adopt several matching and reweighting algorithms to compare similar employees between the treatment and control groups. Table 4.8 presents the ATT results according to six techniques that are often employed in the literature. Specifically, I use entropy balancing (EB), propensity score matching (PSM), Mahalanobis-distance (MD), and inverse-probability-weighted regression adjustment (IPWRA) methods.[27]

The IPWRA estimates satisfy the "double-robustness" property, as only one of the two steps incorporated in the method (that is, the propensity score from the first stage or the outcome regression from the second stage) should be properly specified to yield a consistent treatment effect (Wooldridge, 2007). If both are correctly specified, then the efficiency of the estimate increases. EB is a relatively new reweighting method (Hainmueller, 2012). In the estimation procedure, EB directly encompasses the covariate balance into the weight function that adjusts the control group. Hence, because this entropy maximisation technique orthogonalises the data to the observed characteristics subject to the pre-specified constraints relating to the known covariate moments (e.g., means and variances), there is no need to perform post-match balance checks, as in the case of other matching methods (Hainmueller and Xu, 2013). As a corollary, EB achieves exact balance without any loss of observations.

Interestingly, the ethnic penalties faced by Black, Asian and Mixed/Other groups are very similar to the ones estimated by the OLS regressions across most of the methods discussed above. For example, the estimates from the EB and the IPWRA algorithms for the group experiencing the highest ethnic pay gaps (Black men) is -0.182 (relative to -0.183 from the OLS model). In contrast, the Mahalanobis-distance techniques (columns 4 and 5 of Table 4.8) seem to inflate

[27] *Presenting each method's technical details is outside the scope of this paper, given that matching algorithms are popular in the literature. However, the reader can find further information in the following studies. EB: (Hainmueller, 2012; Hainmueller and Xu, 2013); PSM (Rosenbaum and Rubin, 1983, 1985); MD (Cochran and Rubin, 1973; Carpenter, 1977); IPWRA (Wooldridge, 2007; Imbens and Wooldridge, 2009). I derive the estimates of the matching algorithms using the* kmatch *Stata module developed by Jann (2017b). For each method, I experimented with various bandwidth parameters (e.g., specified caliper and number of nearest neighbours) and functions, but the ATT results did not diverge significantly.*

the ethnic wage gaps across most employee groups. This likely implies that the coefficients on ethnicity are sensitive to the metric used to measure the distance between the treated and control observations in the MD matching (i.e., the Euclidean distance standardised by the covariance matrix of the observed variables). However, for all other techniques, the OLS estimates do not seem particularly sensitive to possible misspecification issues and selection on observables.

Table 4.8. Sensitivity analysis: ATT using different matching and reweighting methods

Ethnic group	Gender	OLS	(1) EB	(2) PSM (KF)	(3) PSM (NN)	(4) MD (KF)	(5) MD (NN)	(6) IPWRA
Black	Women	-0.046**	-0.032	-0.041	-0.039	-0.036	-0.085**	-0.033
		(0.018)	(0.026)	(0.027)	(0.029)	(0.023)	(0.026)	(0.019)
	Men	-0.183***	-0.182***	-0.173***	-0.181***	-0.226***	-0.231***	-0.182***
		(0.029)	(0.037)	(0.040)	(0.040)	(0.036)	(0.040)	(0.029)
	Observations used	47,270	47,270	43,285	2,367	43,728	2,722	44,253
Asian	Women	-0.020	-0.022	-0.013	-0.031	-0.062***	-0.076***	-0.022
		(0.012)	(0.017)	(0.019)	(0.019)	(0.016)	(0.018)	(0.012)
	Men	-0.042***	-0.049*	-0.059**	-0.057**	-0.118***	-0.130***	-0.048***
		(0.014)	(0.020)	(0.022)	(0.022)	(0.019)	(0.021)	(0.014)
	Observations used	47,270	47,270	36,514	7,814	44,694	8,514	45,409
Mixed/ Other	Women	-0.036*	-0.038	-0.030	-0.047	-0.090***	-0.093***	-0.038*
		(0.018)	(0.025)	(0.028)	(0.027)	(0.025)	(0.027)	(0.019)
	Men	-0.013	-0.013	-0.008	0.016	-0.105**	-0.080*	-0.012
		(0.023)	(0.033)	(0.038)	(0.037)	(0.033)	(0.036)	(0.023)
	Observations used	47,270	47,270	22,049	3,323	43,930	3,409	44,684

Notes:
(1): Entropy balancing. Balance is achieved based on the first moments (means).
(2): Propensity-score matching based on the Epanechnikov kernel function (KF). This method uses logistic regression to calculate the propensity score of each employee.
(3): Propensity-score matching based on nearest-neighbour (NN) matching with replacement (five neighbours). This method uses pairs of individuals (from the treatment and control groups) with a difference in their propensity score less than one percentage point, in absolute terms (caliper: 0.01).
(4): Mahalanobis-distance matching based on the Epanechnikov kernel function (KF).
(5): Mahalanobis-distance matching based on the nearest-neighbour (NN) matching with replacement (five neighbours).
(6): Inverse-probability-weighted regression adjustment (IPWRA). This method uses probit regression to calculate each employee's propensity to be treated and subsequently weights the individuals according to the inverse of this probability.
The parentheses show the analytic standard errors (based on influence functions) for the matching methods and the robust standard errors for the OLS models.
The variables used for the matching algorithms are the same as in Table 4.5: Age, Age squared, Region of workplace, Partnered, Any dependent children in family aged <19, Health problems lasting >1 year, Subject area of first degree, High-status universities, Good degree, Managerial/Professional job, Industry sector, Medium/Large enterprises (>=50 employees), Public sector, Part-time work, Permanent job, Years in current employer (Tenure), Total usual weekly hours in main job, Survey year dummies, Survey month dummies.
The variation in observations across specifications reflects the different algorithms used in each matching approach. The reported number of observations corresponds to the total sample size for both genders.
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*
Source: APS 2013-2018

4.5.2 Unobservable selection

This subsection presents the results from the partial identification method suggested by Oster (2017), which measures the coefficient stability under the existence of unobservable characteristics that affect both outcome and treatment. This method explores how the coefficients of the explanatory variable of interest (that is, ethnicity) change according to the relationship between the unobservable factors and ethnicity, relative to the covariance between the observable characteristics and ethnicity. This approach is based on a critical assumption that the selection on observable characteristics is instructive about the selection on omitted (unmeasured) factors. Oster's approach extends the method suggested by Altonji, Elder and Taber (2005), who proposed an estimator (δ) gauging the size of selection on unobservable traits as a proportion of the selection on controlling (observed) factors required to cause a zero-effect of the coefficient of interest (that is, θ=0). In other words, δ determines how large the effect of the unobserved characteristics should be to eliminate the impact of ethnicity on wages (that is, to produce a θ coefficient of zero).

Oster (2017) emphasises that merely looking at the coefficient movements (that is, the changes of θ after including the control variables) is not adequate to evaluate the omitted variable bias. She shows that the ratio of correlations mentioned above (δ) also depends on the maximum R-squared (R_{max}), which, in the current setting, represents the R^2 from a hypothetical regression of log hourly wages on ethnicity and all other observed and unobserved (omitted) variables. By selecting the appropriate values of δ and R_{max}, this method permits calculating consistent estimates of ethnicity's effect on wages. Specifically, Oster proposes that, in empirical research, equal selection (that is, δ=1) is a reasonable upper limit on δ. Moreover, building on a sample of published articles, she finds that a cut-off value of R_{max}=1.3*R^2 (where R^2 refers to the regression with all observed variables) yields a robust (bias-adjusted) estimation of the treatment in nearly half of non-randomised studies and in 90% of randomised settings. She also posits that R_{max} need not be as high as 1 to provide a sufficient bound because of possible measurement errors in the dependent variable.

Following Oster (2017), I evaluate the stability of ethnicity's coefficients based on a range of assumptions about the R_{max} bounds, as shown in the leftmost column of Table 4.9. Negative values of δ imply that the sign of the correlation between ethnicity and regressors is the opposite of the relationship between ethnicity and unobserved variables. Interestingly, the results for Black men reveal that the impact of unobservable characteristics would need to be immense to produce an ethnicity coefficient of zero (Panel A). More specifically, δ needs to range from -22.3 (assuming the extreme case of R_{max}=1) to -134.3 (if R_{max} equals to 1.3 times the R-squared from the OLS regression with the complete set of observable controls) to eliminate the impact of being Black male on wages. For instance, the value of -22.3 means that the degree of selection on unobserved confounders should be 22.3 times that of selection on observables to offset the ethnic differences in wages between White and Black men, as estimated by the OLS models. Because such values seem unrealistic, these findings provide confidence that the ethnic penalties for Black men persist in the UK labour market, and there is firm evidence of racial discrimination.

Given that I have included an extensive and carefully selected set of observed characteristics in the regression analyses that lead to a relatively high R^2 (45% for men and 42% for women), I expect that the real δ should not be very large. A $|\delta|<1$ would imply that the interdependence between the unobserved confounding factors (such as ability, economic motivation, network effects, career aspirations and parental socio-economic background) and the potentially endogenous ethnicity is weaker than the correlation between the observed variables and ethnicity. The results show that Asian men also likely encounter ethnic discrimination in the job market relative to their White counterparts, as the δ value that would completely nullify their pay disadvantage varies from 1.2 to 7.3 for different scenarios of R_{max}. In contrast, the δ values required to reach the same result (θ=0) for Black and Asian women are lower than those for men. The latter figures do not reject the possibility of racial discrimination against Black and Asian females in the labour market. However, there is relatively greater uncertainty about this conclusion in the presence of potential unobserved drivers of productivity.

Moreover, in Panel B of Table 4.9, I choose the recommended value for δ that suggests equal selection (that is, δ=1), so that I can identify the upper bounds for θ, based on different assumptions about R_{max}. On all occasions for Black and Asian men, the identified values of θ do not contain zero, indicating that the OLS results presented in section 4.4.1 are robust to this sensitivity test. In contrast, for men from Mixed/Other backgrounds and women from all ethnic minority groups, the coefficient bounds are broader, meaning that the degree of selection on unobservables needed to cancel out the effect of ethnicity is plausible, particularly when assuming a R_{max} equal to or greater than two times the R^2.

In brief, there is supportive evidence of ethnic discrimination against Black and Asian men in terms of earnings even in presence of selection on unobservables, as the size and the sign of ethnicity's coefficients remain robust after implementing Oster's (2017) partial identification method. It is difficult to explicitly distinguish which form of discrimination discussed in section 4.2 applies in this context. However, the fact that the analysis sample comprises UK-born graduates, many of whom have worked for the same enterprise for several years, suggests that there should be sufficient information available to employers about employees' productivity-related characteristics. Hence, the present work's findings arguably speak in favour of the taste-based discrimination (rather than the statistical discrimination) against Black and Asian men. For women, the picture regarding the mechanisms that trigger wage inequalities is less clear than for men, as the smaller extent of pay gaps amongst females makes is harder to disentangle discrimination from other unobserved wage determinants assertively.

Table 4.9. Sensitivity analysis: Oster's (2017) method

Panel A. Delta (δ) values resulting in a zero-treatment effect ($\delta\|\theta$=0)						
		Men			**Women**	
R_{max}	Black	Asian	Mixed/Other	Black	Asian	Mixed/Other
Min [1.3R^2, 1]	-134.3	7.3	1.1	-2.7	3.5	3.3
Min [1.5R^2, 1]	-70.4	3.8	0.6	-1.6	2.1	1.9
Min [2R^2, 1]	-32.2	1.8	0.3	-0.8	1.0	1.0
1	-22.3	1.2	0.2	-0.6	0.7	0.7
R^2	0.449			0.418		
Panel B. Treatment estimates (θ) when δ equals to 1 ($\theta\|\delta$=1)						
		Men			**Women**	
R_{max}	Black	Asian	Mixed/Other	Black	Asian	Mixed/Other
Min [1.3R^2, 1]	-0.180	-0.037	-0.001	-0.028	-0.014	-0.024
Min [1.5R^2, 1]	-0.179	-0.032	0.011	-0.016	-0.010	-0.017
Min [2R^2, 1]	-0.174	-0.018	0.038	0.012	-0.000	0.002
1	-0.170	-0.008	0.060	0.033	0.008	0.017
R^2	0.449			0.418		

Note: The full set of independent variables (observed controls) used in Oster analysis comprises the variables included in the main OLS regressions of this study. Specifically, these variables are: Ethnicity, Age, Age squared, Region of workplace, Partnered, Any dependent children in family aged <19, Health problems lasting >1 year, Subject area of first degree, High-status universities, Good degree, Managerial/Professional job, Industry sector, Medium/Large enterprises (>=50 employees), Public sector, Part-time work, Permanent job, Years in current employer (Tenure), Total usual weekly hours in main job, Survey year dummies, Survey month dummies.
Source: APS 2013-2018

4.5.3 Further robustness checks

I conducted certain further robustness tests to evaluate the stability of the main OLS regression coefficients and decomposition analysis results.

4.5.3.1 Limiting employees' age to 30-50 years

The first robustness check restricts the sample to prime ages (that is, 30–50-year-old employees). The reason is that there could be unobserved selection effects on the probability of exiting the labour market after 50 (for example, because of inequalities in employment opportunities, health status, the probability of receiving state or private pension, the likelihood of switching to self-employment, cultural reasons, or other circumstances) that might be correlated with ethnicity[28] (Evandrou et al., 2016; Vlachantoni et al., 2017). Moreover, it should take a few years for university graduates to find their job match at the beginning of their working life, whilst there is also limited wage dispersion in the early career phases. As Table 4.A9 in the Appendix shows, the ethnic wage gaps remain remarkably large across genders for the employees aged 30-50 (see columns 1 and 4). In fact, the ethnic penalties for 30–50-year-old Black men (-20.9%) are even higher than those presented in the main OLS regression estimation (Table 4.5), which refer to the life course (-16.7%). The same applies to the rest of the ethnic groups, both for men and women.

Secondly, in columns 2 and 5 of Table 4.A9 in the Appendix, I have dropped the variables relating to occupation and job tenure among employees aged between 30 and 50. The logic behind excluding these variables from the regression is linked to concerns that they could be "bad controls", in the sense that they are endogenous covariates that could themselves be outcomes of ethnicity[29] (Angrist and Pischke, 2009). For instance, if ethnicity determines employees' occupation (if, for example, there is a causal relationship between ethnicity and occupation), then controlling for occupation group would bias ethnicity's effect on wages. This selection bias would imply that ethnicity alters the composition of employees working in managerial/professional occupations. The results show that removing occupation and tenure from the regression

[28] Indeed, the proportion of White workers aged 51-65 covers 18.3% of the total sample of White employees aged 19-65, whereas the corresponding percentage of non-White workers is only 4.8%. In contrast, within the prime ages (30-50), the participation share is identical between both groups (57% versus 59%).

[29] Despite this reasonable concern, many labour market studies (e.g., Tomaskovic-Devey, Thomas and Johnson, 2005; Brynin and Güveli, 2012) include occupation controls in their analysis when estimating ethnicity's effect on wages.

specification increases (by approximately two percentage points) the magnitude of ethnic pay gaps (columns 2 and 5 of Table 4.A9 in the Appendix). This change in ethnic minorities' coefficients probably reflects the findings in descriptive analysis section that ethnic minority employees are less likely to work in higher-status occupations, and they stay in the same company for a shorter time than White graduates. However, the "bad controls" should not be a significant issue in the present context because of the relatively small selection bias size.

Thirdly, including individuals with a higher degree (such as Masters, Doctorate, or "Postgraduate Certificate in Education") in the regression sample of 30–50-year-old workers makes little difference to the effect of ethnicity on wages (columns 3 and 6 of Table 4.A9 in the Appendix). Specifically, the ethnic penalties for Black men, Black women, and Asian men in the extended sample are identical to those for the employees holding a bachelor's degree only. In contrast, the ethnic pay disadvantage for Asian women and employees from Mixed/Other backgrounds (relative to their White counterparts) is 1.5-2 percentage points lower in the sample that also includes higher-degree holders. The average earnings premium of a higher degree relative to a first degree stands at 7.7% for men and 12.4% for women. This is comparable with the premium estimated by Conlon and Patrignani (2011), who drew on the Labour Force Survey data for years 1996-2009 to show that the respective returns to Masters degrees are 9%-10% and to Doctorate degrees 16%-17% (on average for both genders).

4.5.3.2 Other checks

As an additional sensitivity check, I excluded Chinese and people from "any other Asian backgrounds" from the main OLS regression sample (which includes employees aged 19-65). This is because the Chinese may have different characteristics that affect their labour market performance, rendering the Asian group markedly heterogeneous (Mok and Platt, 2018). However, dropping Chinese and "any other Asian" workers from the analysis has a slight effect on the results, given that their proportion in the total sample is small (0.5%, as alluded to in Table 4.1). Specifically, the (un-tabulated) Asian ethnic group's coefficient changes to -0.059 for men (from -0.042 in the original regression) and -0.012 for women (from -0.020) in the models containing the complete list of explanatory variables.

Furthermore, the reduced number of observations may have compromised the statistical significance of the Black ethnic group for women in the separate models by subject of study (Panel B of Table 4.6). To test this, I ran a pooled regression for both genders incorporating the interaction term "*Ethnicity*Male*" in the model. The interaction effect of "*Black*Male*" is negative and statistically significant for the STEM and LEM subjects (at the 5% level) but insignificant for the combined and "other" subjects of study, keeping all else fixed. This is consistent with my main findings, establishing that ethnic penalties are more considerable for Black men than women, even within the same subject area of first degree.

Moreover, the subjects allied to medicine include the non-traditional nursing degrees and many self-employed doctors, such as General Practitioners (Walker and Zhu, 2018). The APS datasets do not capture this type of income, as they contain only employees that report wages during the interview week. Nonetheless, dropping the employees that hold a medical-related degree from the analysis makes little difference to the OLS results.

Also, previous research has found differences in employment and earnings outcomes across various ethno-religious groups in the UK (e.g., Longhi, Nicoletti and Platt, 2013; Khattab and Johnston, 2015). However, including religion as an additional control did not produce statistically significant coefficients and, therefore, I did not incorporate this variable into the regression analyses.

A final robustness check involves adopting an alternative decomposition method proposed by Daymont and Andrisani (1984), which, apart from the explained and unexplained components, includes a third element relating to the interaction between the endowments effect and the unexplained part. Specifically, this threefold decomposition approach allows possible discrepancies in characteristics (that is, the endowments effect) and returns (that is, the magnitude of coefficients) to exist simultaneously between White and non-White employees. Nevertheless, the interaction term's impact in explaining the wage differential between the two ethnic groups was minor and statistically insignificant across both genders. Therefore, in the present study, I adopted the twofold decomposition method suggested by Neumark (1988), as described in subsections 4.3.3.3 and 4.4.3.

4.6 Conclusion

Utilising the most recent data from the Annual Population Survey and concentrating on UK-born graduates, this study has established that specific ethnic minorities face substantial pay differences in the UK labour market compared to equally qualified White people. The wage differences persist even after controlling for many demographic traits, higher education characteristics and job-specific factors. The key conclusion of this paper is that having graduated from higher education does not eliminate ethnic pay inequalities.

For Black men, the labour market experience is way more disappointing, as their wage penalties stand at higher levels (16.7%) than those of women (4.5%). This difference between men and women may emanate from the long-established gender inequality in wages, but it can also reflect varying occupational and educational choices and diverse discrimination levels across genders. Asian men earn 4.1% less than their White counterparts, keeping all else equal, whereas the wage differences for the markedly heterogeneous Mixed/Other ethnic group prove statistically insignificant (-1.3%). The paper's results remain robust to several matching techniques that adjust for the selection on observable characteristics. Ethnic pay inequalities exist even within subgroups of occupations and company types (based on their size) and are aggravated amongst men.

The decomposition analysis reveals that adjusting the non-White employees' observed characteristics to those of their White counterparts would substantially improve wage gaps by nearly 30%. Specifically, the occupational segregation and the shorter job tenure of ethnic minorities relative to White people make up the primary determinants of pay gaps relating to observed factors. It appears that specific structural barriers filter people from different ethnic backgrounds to certain pathways (related to job sectors, occupations, and roles within occupations), resulting in significant raw pay gaps even amongst graduates. Given that some occupation characteristics are themselves outcomes of ethnicity, the policy discussion needs to step back and identify the mechanisms that render these characteristics remarkably skewed against ethnic minority graduates. For example, the under-representation of non-White employees in high-paid jobs may be associated with historical and cultural patterns that channel ethnic minorities to specific professions, social networks, over-education, and racial bias in the labour market (Elliott and Lindley, 2008; Rafferty, 2012).

By constructing a sample confined to UK-born university graduates, I have mitigated specific effects connected with unobservable characteristics that may impinge on ethnic minorities' wages, such as discrepancies in educational profiles, the knowledge of the British labour market, and the proficiency in the host language. Some UK-born ethnic minority people are still likely brought up using first languages other than English in the home setting, but having received school education in the UK should minimise any language disadvantages. Therefore, this work suggests that better education and English fluency are not a panacea for eradicating ethnic penalties. Instead, some characteristics are still valued unequally between White and ethnic minority employees in the UK labour market.

The crucial question is, why do ethnic pay differences stubbornly persist, especially for men? Are labour market prospects undermined by discrimination/racism or do wage inequalities principally reflect other confounding factors not observed in administrative datasets? Data limitations in non-experimental studies, such as the present, preclude disentangling the extent of racial discrimination from other unobserved factors that also shape the level of wages. Notwithstanding, by implementing a coefficient stability method (Oster, 2017) that partially addresses the issue of selection on unobservable characteristics, this paper finds supportive evidence of racial discrimination, particularly against Black and Asian men.

This study primarily controls for human capital characteristics (such as degree class, subject of study, and type of university attended) that are exogenous to the labour market. However, several human capital factors (such as experience and company-specific skills) are the product of the interaction and social negotiation between employers and workers (Tomaskovic-Devey, Thomas and Johnson, 2005). Some employers are likely to discriminate during the recruitment process (Wood et al., 2009), and non-White employees may encounter unequal treatment that undermines their opportunities for training, promotion or transfer to more privileged occupations. Hence, likely discrimination in the labour market against ethnic minorities should aggravate the ethnic wage gaps through the career trajectory.

Due to lack of data, I cannot investigate the effects of the mechanisms mentioned above which may influence the earnings levels. Nonetheless, there is fragmentary evidence that these mechanisms should partially explain why ethnic penalties exacerbate with age across both genders. Specifically, the present study finds no statistically significant ethnic wage differences for females and most male minorities (except Black men) aged below 30. On the contrary, substantial ethnic pay gaps exist for older employees (aged 31-65), relative to their White counterparts, other things fixed. For instance, the average ethnic penalties for Black men double from 10.1% in early career stages to 19.4% for those over 30.

One non-discriminatory explanation for these discrepancies is that some minorities might invest in human capital at lower rates than the White group after being recruited (Lang and Lehmann, 2012), perhaps because of different expectations about the additional skills' value or increased childcare commitments. Furthermore, if ethnic minority employees are less likely to ask for pay rises and promotions due to cultural attitudes or perceived discrimination (Trades Union Congress, 2021), then this should also contribute to their pay disadvantage relative to their White counterparts.

This paper's findings cannot be directly compared with previous works, as the earlier research focusing on UK-born graduates' wages is particularly thin. The closest work by Zwysen and Longhi (2018), who examined British nationals' earnings soon after their graduation, found only small pay differences among ethnic groups. An older study by Heath and Cheung (2006) centred on the UK working-age population using pooled LFS data for years 2001-2004. The authors controlled for many factors, including the highest educational qualification and whether employees are born in the UK. Among other things, they showed that the life-cycle ethnic penalties for the "Black Mixed", "Black Caribbean", and "Black African" groups were 5%, 11% and 21%, respectively, relative to similarly situated White employees. Comparing these figures with the 16.7% average gap for Black men documented in the present study suggests that, nearly two decades later, ethnic inequalities persist in the UK labour market and, perhaps surprisingly, are of similar magnitude even among the highest-educated workforce group.

There are some limitations of the present work. First, although this study exploits many variables that influence productivity and, thus, earnings, it is impossible to control for all wage determinants, which likely differ between ethnic groups. These factors include the level of labour market attachment, economic motivation, network and negotiation effects, parental socio-economic background, and pre-university characteristics relating to individuals' school quality and neighbourhood conditions. Moreover, in the absence of persuasive instruments in the APS datasets, this analysis does not adjust for the selection into paid employment, which might be an element of the wage differential, especially amongst women. Taken together, although this paper establishes the existence of ethnic discrimination for some groups of employees (such as Black and Asian men), it is not feasible to explicitly quantify the magnitude of pure discrimination relative to other unmeasured characteristics.

Second, the broad ethnic disaggregation adopted in this analysis hides ethnic dissimilarities existing within each of the four ethnic groups (White, Black, Asian, Mixed/Other). This should be of particular importance for the Asian ethnic group. For example, Pakistani and Bangladeshi employees experience the lowest raw wages among all ethnic groups (as shown in Table 4.1), whereas Indian and Chinese people exhibit higher earnings levels than the White majority group. This picture likely reflects discrepancies in the social integration of ethnic minorities and discriminatory attitudes against specific ethnic groups, which might be linked to entrenched prejudices rooted back in disparate post-war settlement patterns of immigrants (Craig, 2012; Ehsan, 2018).

Anti-discrimination legislation and policymakers should not only focus on counteracting wage inequalities but also provide incentives to employers toward establishing equitable opportunities within the working environment. To reduce ethnic wage gaps, policy actions should start with removing impediments to the participation of ethnic minority employees in higher-paying jobs and sectors. Specifically, the UK Government might improve equality by extending the pending legislation regarding the mandatory reporting of ethnicity pay gaps to introducing quotas on the minimum proportion of non-White employees in each company (Berson, 2016), mainly aiming to balance the ethnic representation in managerial/professional occupations. In this respect, interventions targeting to tackle the mechanisms leading to ethnic minorities' under-representation in high-salaried occupations might be more effective if applied early in employees' career, given that the likelihood of holding managerial positions should be correlated with being in the job for many years. Non-White people at leadership and senior levels (such as executive positions) could also serve as role models and mentors, thus partly shaping ethnic minorities' career aspirations and their motivation to rise to the top.

Moreover, new policies could reward companies showing good practice with respect to the fair treatment of ethnic minorities (for example, through a deduction in income tax). On employers' side, non-discriminatory companies could diversify the ethnic composition of their recruitment staff, as this would probably increase the likelihood of ethnic minorities being hired and the number of non-White people applying for jobs (Giuliano, Levine and Leonard, 2009). Similarly, implementing monitoring procedures would help employers understand why policies about equal treatment fail in practice.

Decision-makers should also comprehend the reasons for heterogeneity in the ethnic experience between subgroups of employees and across ethnic groups. For instance, the present work shows that Black men see substantially higher earnings gaps within the medium- or large-sized companies than the micro/small enterprises. Similarly, ethnic inequalities worsen with earnings for Black women, as they experience an augmented pay gap at the top quantile of the wage distribution. Conversely, the ethnic penalties for Asian men are substantially higher for those working for small/micro enterprises, having a non-managerial/professional job, and graduating from a non-prestigious higher education institution.

Finally, policymakers could investigate the career trajectories of ethnic minority people, given that pay differences increase with age. The substantially lower ethnic pay disparities among young employees relative to older workers might imply that recent policy and legislation interventions (such as the 2010 Equality Act) have tackled to some extent the ethnic inequalities in the UK labour market. Hence, the strengthening of anti-discrimination legislation over time may have inhibited prejudiced firms from adopting discriminatory behaviours and practices. In a similar context, Storm, Sobolewska and Ford (2017) show that the "social distance" (expressed by the patterns and attitudes toward intermarriage) between the White group and ethnic minorities has declined over the period 1980-2010 in the UK. The authors state that this reduction in the ethnic social distance between generations reflects certain structural mechanisms, such as the rise in ethnic minorities' proportion in the total population (which advances ethnic diversity and interpersonal contacts), the increase in higher education participation rates (which is correlated with a diminution in racial prejudice), and a change in norms and values, which reject blatant racism. For the above reasons, the taste-based discrimination against more recent graduates might have lessened, suggesting that the increased ethnic penalties for older employees established in the present work likely reflect cohort effects (that is, past discrimination encountered by older cohorts).

Contrariwise, a recent review of field experiments conducted in the UK over the period 1969-2017 provides evidence that there is no reduction in the magnitude of discrimination in the labour market for specific non-White minorities, such as Caribbeans and South Asians (Heath and Di Stasio, 2019). The authors assert that the persistence of racial discrimination despite the existing legislation probably mirrors limited implementation and monitoring mechanisms, the absence of financial motivation for firms to comply with the laws in force, and that ethnic parity is not a priority for governments. Although blatant racism has decreased significantly over the last decades in the UK, the authors claim that it is the subtle racism that takes the form of stereotypical beliefs about ethnic minorities' career incentives, and their language and work-pertinent skills, particularly for employees who emigrated from an underdeveloped country. Similarly, previous evidence from the US contradicts the "legacy of past discrimination" assumption (Thomas, Herring and Horton, 1994). To better evaluate how ethnic pay inequalities amongst UK-born graduates evolve over time for recent cohorts of employees, future work using more extended analysis periods or longitudinal data, such as the Understanding Society and the Longitudinal Education Outcomes, is required.

The share of ethnic minorities in the UK population has grown over the last decades through immigration and fertility. Additionally, the overall higher education participation rates have improved significantly, as one young person in two now enters a university. Hence, the ethnic pay inequalities in the labour market documented in the present study have negative consequences for more people. Further work is required to explore how the wage equalisation and full representation of ethnic minority graduates in the labour market would improve the UK's economy and welfare.

References of Chapter 4

Albrecht, J., Björklund, A. and Vroman, S., 2003. Is There a Glass Ceiling in Sweden? *Journal of Labor Economics,* 21(1), pp. 145-177.

Algan, Y., Dustmann, C., Glitz, A. and Manning, A., 2010. The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *Economic Journal,* 120(542), pp. F4-F30.

Altonji, Joseph G., Elder, Todd E. and Taber, Christopher R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy,* 113(1), pp. 151-184.

Angrist, J.D. and Pischke, J.S., 2009. *Mostly harmless econometrics: an empiricist's companion.* Princeton, N.J.: Princeton University Press.

Arrow, K.J., 1972. Some mathematical models of race discrimination in the labor market In: A.H. Pascal, ed. *Racial discrimination in economic life.* Lexington Books, pp. 187-204.

Arulampalam, W., Booth, A.L. and Bryan, M.L., 2007. Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial & Labor Relations Review,* 60(2), pp. 163-186.

Bandiera, O., Barankay, I. and Rasul, I., 2009. Social Connections and Incentives in the Workplace: Evidence From Personnel Data. *Econometrica,* 77(4), pp. 1047-1094.

Barsky, R., Bound, J., Charles, K.K. and Lupton, J.P., 2002. Accounting for the Black–White Wealth Gap. *Journal of the American Statistical Association,* 97(459), pp. 663-673.

Baum, C., 2016. *Quantile regression (Applied Econometrics).* Boston. Available from: http://fmwww.bc.edu/EC-C/S2016/8823/ECON8823.S2016.nn04.slides.pdf [Accessed 20 December 2020].

Becker, G.S., 1971. *The economics of discrimination.* 2nd ed. Chicago: University of Chicago Press.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y., 2018a. *The relative labour market returns to different degrees.* London: IFS (Institute for Fiscal Studies), (DFE-RR787).

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y., 2018b. *The impact of undergraduate degrees on early-career earnings.* London: IFS (Institute for Fiscal Studies), (DFE-RR808).

Berson, C., 2016. Local labor markets and taste-based discrimination. *IZA Journal of Labor Economics,* 5(1), pp. 1-21.

Berthoud, R., 2000. Ethnic employment penalties in Britain. *Journal of Ethnic and Migration Studies,* 26(3), pp. 389-416.

Bertrand, M., Chugh, D. and Mullainathan, S., 2005. Implicit Discrimination. *American Economic Review,* 95(2), pp. 94-98.

Bertrand, M. and Mullainathan, S., 2004. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review,* 94(4), pp. 991-1013.

Blackaby, D., Leslie, D., Murphy, P. and O'Leary, N., 1999. Unemployment Among Britain's Ethnic Minorities. *The Manchester School,* 67(1), pp. 1-20.

Blackaby, D.H., Leslie, D.G., Murphy, P.D. and O' Leary, N.C., 2002. White/Ethnic Minority Earnings and Employment Differentials in Britain: Evidence from the LFS. *Oxford Economic Papers,* 54(2), pp. 270-297.

Blackaby, D.H., Leslie, D.G., Murphy, P.D. and O'Leary, N.C., 1998. The ethnic wage gap and employment differentials in the 1990s: Evidence for Britain. *Economics Letters,* 58(1), pp. 97-103.

Blinder, A.S., 1973. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources,* 8(4), pp. 436-455.

Boliver, V., 2013. How fair is access to more prestigious UK universities? *British Journal of Sociology,* 64(2), pp. 344-364.

Boliver, V., 2016. Exploring Ethnic Inequalities in Admission to Russell Group Universities. *Sociology,* 50(2), pp. 247-267.

Booth, A.L., Francesconi, M. and Frank, J., 2003. A sticky floors model of promotion, pay, and gender. *European Economic Review,* 47(2), pp. 295-322.

Britton, J., Dearden, L., Shephard, N. and Vignoles, A., 2016. How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background. *IDEAS Working Paper Series from RePEc*.

Brynin, M. and Güveli, A., 2012. Understanding the ethnic pay gap in Britain. *Work, Employment & Society,* 26(4), pp. 574-587.

Caliendo, M. and Kopeinig, S., 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. *IDEAS Working Paper Series from RePEc*.

Card, D., Cardoso, A.R. and Kline, P., 2016. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics,* 131(2), pp. 633-686.

Carmichael, F. and Woods, R., 2000. Ethnic Penalties in Unemployment and Occupational Attainment: Evidence for Britain. *International Review of Applied Economics,* 14(1), pp. 71-98.

Carpenter, R.G., 1977. Matching when covariables are normally distributed. *Biometrika,* 64(2), pp. 299-307.

Cheng, Y. and Heath, A., 1993. Ethnic Origins and Class Destinations. *Oxford Review of Education,* 19(2), pp. 151-165.

Chevalier, A., 2011. Subject Choice and Earnings of UK Graduates. *Economics of Education Review,* 30(6), pp. 1187-1201.

Chiswick, B.R., 1978. The Effect of Americanization on the Earnings of Foreign--born Men. *Journal of Political Economy,* 86(5), pp. 897-922.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A., 2013. Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 176(2), pp. 431-457.

Christopoulou, R. and Monastiriotis, V., 2014. The Greek Public Sector Wage Premium before the Crisis: Size, Selection and Relative Valuation of Characteristics. *British Journal of Industrial Relations,* 52(3), pp. 579-602.

Clark, K., 2015. *Ethnic minority self-employment.* IZA World of Labor, (2015: 120).

Clark, K. and Drinkwater, S., 2002. Enclaves, neighbourhood effects and employment outcomes: Ethnic minorities in England and Wales. *Journal of Population Economics,* 15(1), pp. 5-29.

Clark, K. and Drinkwater, S., 2008. The labour-market performance of recent migrants. *Oxford Review of Economic Policy,* 24(3), pp. 496-517.

Clark, K. and Lindley, J., 2008. Immigrant assimilation pre and post labour market entry: evidence from the UK Labour Force Survey. *Journal of Population Economics,* 22(1), pp. 175-198.

Cochran, W.G. and Rubin, D.B., 1973. Controlling Bias in Observational Studies: A Review. *Sankhya. Series A,* 35(4), pp. 417-446.

Coffield, F. and Vignoles, A., 1997. *Widening participation in higher education by ethnic minorities, women and alternative Students, Report 5, Higher Education in the Learning Society.* London: NCIHE (National Committee of Inquiry into Higher Education).

Cohen, P.N. and Huffman, M.L., 2007. Black Under-representation in Management across U.S. Labor Markets. *The Annals of the American Academy of Political and Social Science,* 609(1), pp. 181-199.

Conlon, G. and Patrignani, P., 2011. *The returns to higher education qualifications.* London: BIS (Department for Business Innovation and Skills), (BIS Research paper number 45).

Craig, G., 2012. The history and pattern of settlement of the UK's black and minority ethnic population. In: G. Craig, K. Atkin, S. Chattoo and R. Flynn, eds. *Understanding 'Race'and Ethnicity: Theory, History, Policy, Practice.* Bristol: The Policy Press, pp. 41-69.

Crawford, C., Dearden, L., Mesnard, A., Sianesi, B. and Shaw, J., 2010. *Ethnic parity in labour market outcomes for benefit claimants in Great Britain.* London: U.C.L. Department of Quantitative Social Science - UCL Institute of Education, (Paper No. 10-07).

Dale, A., Shaheen, N., Kalra, V. and Fieldhouse, E., 2002. Routes into education and employment for young Pakistani and Bangladeshi women in the UK. *Ethnic and Racial Studies,* 25(6), pp. 942-968.

Dale, S.B. and Krueger, A.B., 2014. Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data. *Journal of Human Resources,* 49(2), pp. 323-358.

Darity Jr, W. and Mason, P.L., 1998. Evidence on Discrimination in Employment: Codes of Color, Codes of Gender. *The Journal of Economic Perspectives,* 12(2), pp. 63-90.

Daymont, T.N. and Andrisani, P.J., 1984. Job Preferences, College Major, and the Gender Gap in Earnings. *The Journal of Human Resources,* 19(3), pp. 408-428.

Deloitte UK, 2019. *Ethnicity Pay Report.* London. Available from: https://www2.deloitte.com/uk/en/pages/growth/articles/ethnicity-pay-gap-report.html [Accessed 20 December 2020].

Devine, P.G., 1989. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology,* 56(1), pp. 5-18.

DfE, 2019. *Participation Rates In Higher Education: Academic Years 2006/2007 – 2017/2018 (Provisional).* London: DfE (Department for Education).

Dustmann, C. and Fabbri, F., 2003. Language proficiency and labour market performance of immigrants in the UK. *Economic Journal,* 113(489), pp. 695-717.

Dustmann, C. and Fabbri, F., 2005. Immigrants in the British Labour Market*. *Fiscal Studies,* 26(4), pp. 423-470.

Dustmann, C., Glitz, A., Schönberg, U. and Brücker, H., 2016. Referral-based Job Search Networks. *The Review of Economic Studies,* 83(2), p. 514.

Ehsan, R., 2018. Poll finds more favourable attitudes towards Indian immigrants than Pakistanis and Bangladeshis – why? *The Conversation,* [Online]. Available from: https://theconversation.com/poll-finds-more-favourable-attitudes-towards-indian-immigrants-than-pakistanis-and-bangladeshis-why-96040 [Accessed 20 December 2020].

Elder, T.E., Goddeeris, J.H. and Haider, S.J., 2010. Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Economics,* 17(1), pp. 284-290.

Elliott, R.J.R. and Lindley, J.K., 2008. Immigrant wage differentials, ethnicity and occupational segregation. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 171(3), pp. 645-671.

Equality Act, 2010. c. 15. *Retrieved from the UK National Archives website: http://www.legislation.gov.uk/ukpga/2010/15/contents.* London: HMSO.

Evandrou, M., Falkingham, J., Feng, Z. and Vlachantoni, A., 2016. Ethnic inequalities in limiting health and self-reported health in later life revisited. *Journal of Epidemiology and Community Health,* 70(7), pp. 653-662.

Feng, A. and Graetz, G., 2017. A question of degree: The effects of degree class on labor market outcomes. *Economics of Education Review,* 61, pp. 140-161.

Firpo, S., Fortin, N.M. and Lemieux, T., 2009. Unconditional Quantile Regressions. *Econometrica,* 77(3), pp. 953-973.

Frattini, T., 2017. Evaluating the Labour Market Integration of New Immigrants in the UK. *Social Policy and Society,* 16(4), pp. 645-658.

Friedberg, R.M., 2000. You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics,* 18(2), pp. 221-251.

Gangl, M. and Ziefle, A., 2009. Motherhood, labor force behavior, and women's careers: an empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. *Demography,* 46(2), pp. 341-369.

Giuliano, L., Levine, David I. and Leonard, J., 2009. Manager Race and the Race of New Hires. *Journal of Labor Economics,* 27(4), pp. 589-631.

Gronau, R., 1988. Sex-Related Wage Differentials and Women's Interrupted Labor Careers-the Chicken or the Egg. *Journal of Labor Economics,* 6(3), pp. 277-301.

Guryan, J. and Charles, K.K., 2013. Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. *The Economic Journal,* 123(572), pp. F417-F432.

Hainmueller, J., 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis,* 20(1), pp. 25-46.

Hainmueller, J. and Xu, Y., 2013. ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software,* 54(7), pp. 1-18.

Heath, A. and Cheung, S.Y., 2006. *Ethnic penalties in the labour market: Employers and discrimination.* London: Department for Work and Pensions, (Research Report No 341).

Heath, A.F. and Di Stasio, V., 2019. Racial discrimination in Britain, 1969–2017: a meta-analysis of field experiments on racial discrimination in the British labour market. *The British Journal of Sociology,* 70(5), pp. 1774-1798.

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. *Econometrica,* 47(1), pp. 153-161.

Hernandez, M., Avery, D.R., Volpone, S.D. and Kaiser, C.R., 2019. Bargaining While Black: The Role of Race in Salary Negotiations. *Journal of Applied Psychology,* 104(4), pp. 581-592.

Huber, M., Lechner, M. and Wunsch, C., 2013. The performance of estimators based on the propensity score. *Journal of Econometrics,* 175(1), pp. 1-21.

Imbens, G.W. and Wooldridge, J.M., 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature,* 47(1), pp. 5-86.

Jann, B., 2008. The Blinder–Oaxaca decomposition for linear regression models. *Stata Journal,* 8(4), pp. 453-479.

Jann, B., 2017a. *Why propensity scores should be used for matching.* Available from: https://EconPapers.repec.org/RePEc:boc:dsug17:01 [Accessed 20 December 2020].

Jann, B., 2017b. *kmatch: Stata module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment.* Boston College Department of Economics, (Statistical Software Components S458346). Available from: https://ideas.repec.org/c/boc/bocode/s458346.html [Accessed 20 December 2020].

Jurajda, Š. and Paligorova, T., 2009. Czech female managers and their wages. *Labour Economics,* 16(3), pp. 342-351.

Khattab, N. and Johnston, R., 2013. Ethnic and Religious Penalties in a Changing British Labour Market from 2002 to 2010: The Case of Unemployment. *Environment and Planning A,* 45(6), pp. 1358-1371.

Khattab, N. and Johnston, R., 2015. Ethno-religious identities and persisting penalties in the UK labor market. *The Social Science Journal,* 52(4), pp. 490-502.

Koenker, R., 2005. *Quantile Regression.* Cambridge: Cambridge University Press.

Koenker, R. and Bassett, G., 1978. Regression Quantiles. *Econometrica,* 46(1), pp. 33-50.

Lang, K. and Lehmann, J.-Y.K., 2012. Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature,* 50(4), pp. 959-1006.

Lemos, S., 2013. Immigrant economic assimilation: Evidence from UK longitudinal data between 1978 and 2006. *Labour Economics,* 24, pp. 339-353.

Li, Y., 2018. Unequal Returns: Higher Education and Access to the Salariat by Ethnic Groups in the UK. In: J. Arday and H.S. Mirza, eds. *Dismantling Race in Higher Education.* Basingstoke, Hampshire: Palgrave Macmillan, pp. 103-124.

Li, Y. and Heath, A., 2010. Struggling onto the ladder, climbing the rungs: employment status and class position by minority ethnic groups in Britain (1972-2005). In: J. Stillwell, P. Norman, C. Thomas and P. Surridge, eds. *Population, Employment, Health and Well-being.* Springer.

Li, Y. and Heath, A., 2018. Persisting disadvantages: a study of labour market dynamics of ethnic unemployment and earnings in the UK (2009–2015). *Journal of Ethnic and Migration Studies*, pp. 1-22.

Lindley, J., 2002a. Race or religion? The impact of religion on the employment and earnings of Britain's ethnic communities. *Journal of Ethnic and Migration Studies,* 28(3), pp. 427-442.

Lindley, J., 2002b. The English Language Fluency And Earnings Of Ethnic Minorities In Britain. *Scottish Journal of Political Economy,* 49(4), pp. 467-487.

Lindley, J., Dale, A. and Dex, S., 2004. Ethnic differences in women's demographic, family characteristics and economic activity profiles, 1992 to 2002. *Labour Market Trends,* 112(4), pp. 153-165.

Lindley, J.K., 2016. Lousy pay with lousy conditions: the role of occupational desegregation in explaining the UK gender pay and work intensity gaps. *Oxford Economic Papers,* 68(1), pp. 152-173.

Longhi, S. and Brynin, M., 2017. *The ethnicity pay gap.* Manchester: Equality and Human Rights Commission, (Research report: 108).

Longhi, S., Nicoletti, C. and Platt, L., 2013. Explained and unexplained wage gaps across the main ethno-religious groups in Great Britain. *Oxford Economic Papers,* 65(2), pp. 471-493.

Longhi, S. and Platt, L., 2008. *Pay Gaps Across Equalities Areas.* Equality and Human Rights Commission, (Research report: 9).

Machin, S. and Puhani, P.A., 2003. Subject of degree and the gender wage differential: evidence from the UK and Germany. *Economics Letters,* 79(3), pp. 393-400.

Manning, A. and Swaffield, J., 2008. The gender gap in early-career wage growth. *The Economic Journal,* 118(530), p. 983.

McGregor-Smith, R., 2017. *Race in the workplace: The McGregor-Smith Review.* London: Department for Business Energy & Industrial Strategy.

Mincer, J., 1974. *Schooling, Experience, and Earnings.* New York: National Bureau of Economic Research.

Miranda, A. and Zhu, Y., 2013. English deficiency and the native–immigrant wage gap. *Economics Letters,* 118(1), pp. 38-41.

Modood, T. and Khattab, N., 2016. Explaining Ethnic Differences: Can Ethnic Minority Strategies Reduce the Effects of Ethnic Penalties? *Sociology,* 50(2), pp. 231-246.

Mok, T.M. and Platt, L., 2018. All look the same? Diversity of labour market outcomes of Chinese ethnic group populations in the UK. *Journal of Ethnic and Migration Studies,* 46(1), pp. 87-107.

Naylor, R., Smith, J. and Telhaj, S., 2016. Graduate returns, degree class premia and higher education expansion in the UK. *Oxford Economic Papers,* 68(2), pp. 525-545.

NCIHE, 1997. *Higher Education in the Learning Society (The 'Dearing Report').* London: NCIHE (The National Committee of Inquiry into Higher Education). Available from: http://www.leeds.ac.uk/educol/ncihe/ [Accessed 20 December 2020].

Neal, D., 2004. The measured black-white wage gap among women is too small. *Journal of Political Economy,* 112(1), pp. 1-28.

Neal, D. and Johnson, W., 1996. The role of premarket factors in black-white wage differences. *Journal of Political Economy,* 104(5), pp. 869-895.

Neumark, D., 1988. Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *The Journal of Human Resources,* 23(3), pp. 279-295.

Neumark, D., 2018. Experimental Research on Labor Market Discrimination. *Journal of Economic Literature,* 56(3), pp. 799-866.

Niederle, M. and Vesterlund, L., 2007. Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics,* 122(3), pp. 1067-1101.

Oaxaca, R.L., 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review,* 14(3), pp. 693-709.

Oaxaca, R.L. and Ransom, M.R., 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics,* 61(1), pp. 5-21.

ONS, 2020. Retail Prices Index: Long run series: 1947 to 2019. *Office for National Statistics* [Online]. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/cdko/mm23 [Accessed 20 December 2020].

Oster, E., 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics,* 37(2), pp. 187-204.

Pager, D., Western, B. and Bonikowski, B., 2009. Discrimination in a low-wage labor market: A field experiment. *American Sociological Review,* 74(5), pp. 777-799.

Phelps, E.S., 1972. The Statistical Theory of Racism and Sexism. *The American Economic Review,* 62(4), pp. 659-661.

Phillips, D., 1998. Black Minority Ethnic Concentration, Segregation and Dispersal in Britain. *Urban Studies,* 35(10), pp. 1681-1702.

Platt, L., 2007. Making Education Count: The Effects of Ethnicity and Qualifications on Intergenerational Social Class Mobility. *The Sociological Review,* 55(3), pp. 485-508.

Price, S.W., 2001. The employment adjustment of male immigrants in England. *Journal of Population Economics,* 14(1), pp. 193-220.

Race Relations Act, 1976. c. 74. *Retrieved from the UK National Archives website:* http://www.legislation.gov.uk/ukpga/1976/74/section/4/enacted. London: HMSO.

Rafferty, A., 2012. Ethnic penalties in graduate level over-education, unemployment and wages: evidence from Britain. *Work, Employment & Society,* 26(6), pp. 987-1006.

Reimers, C., 1983. Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics*, p. 570.

Rosenbaum, P.R. and Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika,* 70(1), pp. 41-55.

Rosenbaum, P.R. and Rubin, D.B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician,* 39, pp. 33-39.

Russell Group, 2019. *Supporting under-represented students to access and succeed in higher education.* Available from: https://russellgroup.ac.uk/media/5786/supporting-under-represented-students-to-access-and-succeed-in-he-july-2019.pdf [Accessed 20 December 2020].

Schaafsma, J. and Sweetman, A., 2001. Immigrant earnings: age at immigration matters. *Canadian Journal of Economics/Revue Canadienne d`Economique,* 34(4), pp. 1066-1099.

See, B.H., Torgerson, C., Gorard, S., Ainsworth, H., Low, G. and Wright, K., 2011. Factors that promote high post-16 participation of some minority ethnic groups in England: a systematic review of the UK-based literature. *Research in Post-Compulsory Education,* 16(1), pp. 85-100.

Shields, M.A. and Price, S.W., 1998. The earnings of male immigrants in England: evidence from the quarterly LFS. *Applied Economics,* 30(9), pp. 1157-1168.

Shiner, M. and Modood, T., 2002. Help or Hindrance? Higher Education and the Route to Ethnic Equality. *British Journal of Sociology of Education,* 23(2), pp. 209-232.

Spence, M., 1973. Job Market Signaling. *The Quarterly Journal of Economics,* 87(3), pp. 355-374.

Storm, I., Sobolewska, M. and Ford, R., 2017. Is ethnic prejudice declining in Britain? Change in social distance attitudes among ethnic majority and minority Britons. *The British Journal of Sociology,* 68(3), pp. 410-434.

Swinford, S., 2018. All large employers will be forced to publish their 'ethnicity pay gap' under Government plans. *The Telegraph* [Online]. Available from: https://www.telegraph.co.uk/politics/2018/10/10/large-employers-will-forced-publish-ethnicity-pay-gap-government/ [Accessed 20 December 2020].

Thijssen, L., 2016. *Taste-based versus Statistical Discrimination: Placing the Debate into Context.* The European Commission (GEMM Project), (Deliverable D3. 2).

Thomas, M.E., Herring, C. and Horton, H.D., 1994. Discrimination over the Life Course: A Synthetic Cohort Analysis of Earnings Differences between Black and White Males, 1940-1990. *Social Problems,* 41(4), pp. 608-628.

Times Higher Education, 2014. REF 2014: Overall Ranking of Institutions. *Times Higher Education* [Online]. Available from: https://www.timeshighereducation.com/sites/default/files/Attachments/2014/12/17/k/a/s/over-14-01.pdf [Accessed 20 December 2020].

Tomaskovic-Devey, D., Thomas, M. and Johnson, K., 2005. Race and the Accumulation of Human Capital across the Career: A Theoretical Model and Fixed-Effects Application. *American Journal of Sociology,* 111(1), pp. 58-89.

Topa, G., 2011. Labor Markets and Referrals. In: J. Benhabib, A. Bisin and M.O. Jackson, eds. *Handbook of Social Economics.* North-Holland, pp. 1193-1221.

Topham, G., 2018. £3.2bn UK pay gap for black, Asian and ethnic minority workers. *The Guardian* [Online]. Available from: https://www.theguardian.com/money/2018/dec/27/uk-black-and-ethnic-minorities-lose-32bn-a-year-in-pay-gap [Accessed 20 December 2020].

Trades Union Congress, 2021. *One in three BME workers say they have been unfairly turned down for a job: TUC poll* [Online]. London: TUC (Trades Union Congress). Available from: https://www.tuc.org.uk/news/one-three-bme-workers-say-they-have-been-unfairly-turned-down-job-tuc-poll [Accessed 30 March 2021].

Valbuena, J. and Zhu, Y., 2018. The returns to UK degrees for foreign-educated graduates. *British Educational Research Journal,* 44(3), pp. 440-462.

Viitanen, T., 2012. The motherhood wage gap in the UK over the life cycle. *Review of Economics of the Household,* 12(2), pp. 259-276.

Vlachantoni, A., Feng, Z., Evandrou, M. and Falkingham, J., 2017. Ethnic elders and pension protection in the United Kingdom. *Ageing and Society,* 37(5), pp. 1025-1049.

Vlassopoulos, M. and Siddique, A., 2020. Competitive Preferences and Ethnicity: Experimental Evidence from Bangladesh. *The Economic Journal,* 130(627), pp. 793-821.

Waldfogel, J., 1998. Understanding the 'Family Gap' in Pay for Women with Children. *Journal of Economic Perspectives,* 12(1), pp. 137-156.

Walker, I. and Zhu, Y., 2011. Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales. *Economics of Education Review,* 30(6), pp. 1177-1186.

Walker, I. and Zhu, Y., 2013. *The impact of university degrees on the lifecycle of earnings: some further analysis.* London: BIS (Department for Business Innovation and Skills).

Walker, I. and Zhu, Y., 2018. University selectivity and the relative returns to higher education: Evidence from the UK. *Labour Economics,* 53, pp. 230-249.

Wolfolds, S.E. and Siegel, J., 2019. Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal,* 40(3), pp. 432-462.

Wood, M., Hales, J., Purdon, S., Sejersen, T. and Hayllar, O., 2009. *A test for racial discrimination in recruitment practice in British cities.* London: Department for Work and Pensions, (Research Report No 607).

Wooldridge, J.M., 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics,* 141(2), pp. 1281-1301.

Yao, Y. and van Ours, J.C., 2015. Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics,* 34(C), pp. 76-85.

Yap, M. and Konrad, A.M., 2009. Gender and Racial Differentials in Promotions: Is There a Sticky Floor, a Mid-Level Bottleneck, or a Glass Ceiling? *Relations Industrielles,* 64(4), pp. 593-619.

Zschirnt, E. and Ruedin, D., 2016. Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies,* 42(7), pp. 1115-1134.

Zuccotti, C.V., 2015. Do Parents Matter? Revisiting Ethnic Penalties in Occupation among Second Generation Ethnic Minorities in England and Wales. *Sociology,* 49(2), pp. 229-251.

Zwysen, W. and Longhi, S., 2018. Employment and earning differences in the early career of ethnic minority British graduates: the importance of university career, parental background and area characteristics. *Journal of Ethnic and Migration Studies,* 44(1), pp. 154-172.

Appendix of Chapter 4

Table 4.A1. Description of the variables used in this paper

Variable	Values	Description – Notes
Age	19-65 years	Age of respondent (employee) at the interview time.
Any dependent children in family aged <19	Yes, No	A binary variable denoting whether the employee has any dependent children aged under 19, including all those aged 0-15 and students aged 16-18 in full-time education.
Class of first degree	First class, Upper second class, Third class, Pass, Other/Unknown.	This classification is presented in Table 4.3 only. In the regression models, I use the “good degree” variable instead (see below), because of sample size limitations.
Ethnic group	White, Black, Asian, Other/Mixed	Employee’s self-reported ethnicity. I adopt the recommended broader classification based on the 2011 National Statistics, except that “Other” and “Mixed” ethnic groups are combined into one category (due to the small number of observations). The White category includes UK-born employees from White British, Irish, Gypsy or Irish Traveller, and any other White backgrounds, thus containing some White ethnic minorities. However, in line with the practice implemented in other studies in the field, the “ethnic minority” term used throughout this work considers only non-White individuals. The Black ethnic group comprises people from African, Caribbean or any other Black background. The Asian ethnic group includes Indian, Pakistani, Bangladeshi, Chinese and people from any other Asian background. The “Other/Mixed” category contains employees from Mixed/Multiple ethnic backgrounds (such as White and Black Caribbean, White and Black African, White and Asian, or any other mixed/multiple ethnic groups), Arab minorities, and any other group not mentioned in any of the rest categories.
Good degree	Yes, No	A binary variable capturing employees who attained either a first-class or an upper second-class honours degree.
Health problems lasting >1 year	Yes, No	A binary variable showing whether an employee has long-term health conditions/illnesses, which do not, however, act as a deterrent to continuing working.
High-status universities	Yes, No	A binary variable capturing employees who graduated from a prestigious, research-intensive university. High-status universities cover 41 institutions, including the Russell Group (RG) universities and any higher education institutions that demonstrated a higher ranking than the lowest RG according to the 2014 Research Excellence Framework (REF). Rankings are based on the grade point average (GPA) and take into account the output, impact and environment profiles of universities (Times Higher Education, 2014). The top 41 higher education institutions are the following: Imperial College London, London School of Economics and Political Science, University of Oxford, University of Cambridge, Cardiff University, King’s College London, University College London, University of Warwick, London School of Hygiene and Tropical Medicine (University of London), University of Edinburgh, University of Bristol, Queen Mary University of London, University of Sheffield, University of York, University of Bath, University of Manchester, University of Southampton, Lancaster University, Durham University, University of Leeds, University of St. Andrews, University of East Anglia, University of Glasgow, Liverpool School of Tropical Medicine, University of Nottingham, Newcastle University, Royal Holloway (University of London), Swansea University, University of Exeter, University of Birmingham, Cranfield University, University of Liverpool, Heriot-Watt University, University of Essex, Aston University, University of Strathclyde, University of Reading, University of Dundee, University of Sussex, Cardiff Metropolitan University, Queen’s University Belfast.

Continued on next page

Table 4.A1. (continued)		
Industry sector (grouped)	Public Admin/Education/Health, Banking/Finance, Trade/Hotel/Restaurants, Transport/Communication, Manufacturing/Construction, Other industry	Represents the employee's industry group in the main job. The "Other industry" category comprises the "Agriculture, forestry & fishing", "Energy & water" and "Other services" sectors.
Male	Yes, No	A binary variable capturing male employees.
Managerial/Professional job	Yes, No	This dummy variable represents the employees working in managerial or professional occupations, which include "Managers, directors & senior officials", "Professional occupations" and "Associate professional & technical occupations". The non-managerial/professional jobs comprise "Administrative & secretarial", "Skilled trades", "Caring, leisure & other services", "Sales & customer service", "Process, plant & machine operatives", and "Elementary" occupations.
Part-time work	Yes, No	A binary variable designating whether the employee is working on a part-time basis.
Partnered	Yes, No	This variable captures those employees who are married, cohabiting or in a civil partnership.
Permanent job	Yes, No	A binary variable denoting whether the employee is working permanently or not (e.g., seasonal work, fixed-time contract).
Public sector	Yes, No	A binary variable showing whether the respondent is employed in the public sector.
Region of workplace	London, South East, Northern, Rest of England, Wales, Scotland, Northern Ireland.	This variable represents the region of the workplace according to the standard Government Office Regions. I followed this broader region grouping, as no statistically significant differences were found in the model results when applying a more detailed classification (especially for the "Rest of England" category). The "Northern" region includes Greater Manchester, Merseyside, Rest of North West, Tyne & Wear, and Rest of North East. The "Rest of England" category groups together all other English regions (South Yorkshire, West Yorkshire, Rest of Yorkshire & Humberside, East Midlands, East Anglia, South West, and West Midlands).
Subject area of first degree	Health, Sciences, Engineering/Technology, Social studies, Law/Business/Finance, Arts/Humanities/Education, Combined subject	The subject area of the first-degree course based on the principal codes of the "Joint Academic Coding System" (JACS) classification. I have applied this broader grouping (7 categories) because of data limitations. Each category consists of the following subject areas: 1."Health": Medicine and dentistry, Subjects allied to Medicine; 2."Sciences": Biological sciences, Veterinary science, Agricultural sciences, Physical/Environmental sciences, Mathematical sciences, Computer sciences; 3."Engineering/Technology": Engineering, Technologies, Architecture & related studies; 4."Social studies": Social studies (e.g., Economics, Politics, Sociology, Social Policy); 5."Law/Business/Finance": Law, Business & administrative studies; 6."Arts/Humanities/Education": Mass communications & documentation, Languages, Historical & philosophical studies, Creative arts & design, Education; 7."Combined subjects" include joint degrees in one or over one area of subject (e.g., History & Politics, Economics & Mathematics).

Continued on next page

Table 4.A1. (continued)

Subject area of first degree (broader groups)	STEM, LEM, Other subject, Combined subject	This grouping is used only in the section of the regression analysis exploring heterogeneous effects of ethnicity on wages (Table 4.6). Following relevant studies (see, for instance, Walker and Zhu, 2011; Britton et al., 2016), I have further grouped the subject codes of first degree, as defined by the JACS classification, into 4 broader categories: STEM, LEM, Other, and Combined subject. The STEM category comprises subjects related to Sciences (Biological, Agricultural and Physical Sciences), Technology, Engineering and Mathematics. This group also includes Architects and Health-related subjects, which, however, make up a small part of graduates' sample and barely affect the overall results. LEM denotes subjects that lie in the areas of Law, Economics and Management (i.e., Social studies, Business and Financial studies, etc.). The "Other" category is formed by all the remaining first-degree subjects (Languages, Humanities, Arts, Education, Mass Communication, etc.). The "Combined" category refers to degrees joint in one or over one subject areas.
Total usual weekly hours in main job	1-97	This variable denotes the employee's average number of working hours per week (including overtime) in the main occupation.
Workplace size	Micro/Small enterprise (<50 employees), Medium/Large enterprise (>=50 employees)	This variable represents the company size based on the total number of workers at the employee's main place of work.
Years in current employer (Tenure)	0-48	This variable represents the number of years working for the current employer.
Year obtained first degree / Years after graduation	-	These variables are used only in the Appendix tables (descriptive statistics) to provide a more detailed picture of the graduates' (employees') profile. The number of years that have passed since graduation are calculated at the interview time. Because of the large number of missing cases, I do not include these two variables in the regression analysis.

Source: APS 2013-2018

Table 4.A2. Mean characteristics: White versus non-White employees

Variable	**Non-White**	**White**	**Difference**
Log real hourly wage (in December 2018 constant prices)	2.751	2.813	-0.06***
Real hourly wage (£) (in December 2018 constant prices)	17.669	18.925	-1.26***
Demographic characteristics			
Male	0.444	0.452	-0.01
Age	33.933	38.970	-5.04***
London	0.379	0.111	0.27***
South East	0.159	0.151	0.01
Northern	0.138	0.187	-0.05***
Rest of England	0.272	0.305	-0.03***
Wales	0.016	0.095	-0.08***
Scotland	0.034	0.132	-0.10***
Partnered	0.543	0.716	-0.17***
Any dependent children in family aged <19	0.468	0.432	0.04***
Health problems lasting >1 year	0.183	0.237	-0.05***
Higher education characteristics			
Year obtained first degree	2004.780	2000.694	4.09***
Years after graduation	11.020	15.025	-4.00***
Russell Group universities	0.205	0.238	-0.03***
High status universities	0.262	0.335	-0.07***
Good degree	0.562	0.571	-0.01
Health	0.089	0.109	-0.02***
Sciences	0.196	0.194	0.00
Engineering/Technology	0.058	0.091	-0.03***
Social studies	0.103	0.086	0.02***
Law/Business/Finance	0.258	0.146	0.11***
Arts/Humanities/Education	0.147	0.232	-0.08***
Combined subject	0.148	0.142	0.01
Occupation/Sector characteristics			
Managerial/Professional job	0.699	0.762	-0.06***
Public Admin/Education/Health	0.364	0.448	-0.08***
Banking/Finance	0.275	0.198	0.08***
Trade/Hotel/Restaurant	0.136	0.092	0.04***
Transport/Communication	0.106	0.087	0.02***
Manufacturing/Construction	0.071	0.108	-0.04***
Other industry	0.048	0.066	-0.02***
Micro/Small enterprises (<50 employees)	0.346	0.387	-0.04***
Medium/Large enterprises (>=50 employees)	0.654	0.613	0.04***
Public sector	0.307	0.390	-0.08***
Part-time work	0.168	0.189	-0.02***
Permanent job	0.930	0.946	-0.02***
Years in current employer (Tenure)	5.102	7.797	-2.70***
Total usual weekly hours in main job	37.741	38.764	-1.02***
Observations	**2,921**	**46,735**	

Note: I examine the equality of wage means between the two groups using standard Student's t-tests and proportion tests.
"N. Ireland" variable is not reported due to the underlying small cell size.
** p < 0.1, ** p < 0.05, *** p < 0.01.*
Source: APS 2013-2018

Table 4.A3. Mean characteristics by gender

Variable	**Men**	**Women**	**Difference**
Log real hourly wage (in December 2018 constant prices)	2.928	2.712	-0.22***
Real hourly wage (£) (in December 2018 constant prices)	21.435	16.728	-4.71***
Demographic characteristics			
Non-White	0.058	0.060	0.00
Age	39.254	38.198	-1.06***
London	0.149	0.109	-0.04***
South East	0.159	0.146	-0.01***
Northern	0.178	0.189	0.01***
Rest of England	0.301	0.304	0.00
Wales	0.080	0.099	0.02***
Scotland	0.116	0.134	0.02***
N. Ireland	0.017	0.019	0.00*
Partnered	0.723	0.692	-0.03***
Any dependent children in family aged <19	0.419	0.446	0.03***
Health problems lasting >1 year	0.221	0.245	0.02***
Higher education characteristics			
Year obtained first degree	1999.639	2002.042	2.40***
Years after graduation	16.067	13.694	-2.37***
Russell Group universities	0.253	0.222	-0.03***
High status universities	0.352	0.314	-0.04***
Good degree	0.534	0.600	0.07***
Health	0.040	0.163	0.12***
Sciences	0.249	0.149	-0.10***
Engineering/Technology	0.169	0.023	-0.15***
Social studies	0.078	0.094	0.02***
Law/Business/Finance	0.156	0.150	-0.01*
Arts/Humanities/Education	0.174	0.270	0.10***
Combined subject	0.134	0.150	0.02***
Occupation/Sector characteristics			
Managerial/Professional job	0.804	0.722	-0.08***
Public Admin/Education/Health	0.276	0.581	0.31***
Banking/Finance	0.248	0.166	-0.08***
Trade/Hotel/Restaurant	0.098	0.093	-0.01**
Transport/Communication	0.138	0.047	-0.09***
Manufacturing/Construction	0.162	0.060	-0.10***
Other industry	0.079	0.054	-0.02***
Micro/Small enterprises (<50 employees)	0.357	0.408	0.05***
Medium/Large enterprises (>=50 employees)	0.643	0.592	-0.05***
Public sector	0.257	0.491	0.23***
Part-time work	0.059	0.294	0.24***
Permanent job	0.958	0.935	-0.02***
Years in current employer (Tenure)	7.919	7.407	-0.51***
Total usual weekly hours in main job	41.984	36.006	-5.98***
Observations	**22,420**	**27,261**	

Note: I examine the equality of wage means between the two groups using standard Student's t-tests and proportion tests.
** p < 0.1, ** p < 0.05, *** p < 0.01.*
Source: APS 2013-2018

Table 4.A4. Frequencies and mean log(wage) by earnings determinants: White versus non-White employees

Variable	Observations		Mean log (wage)		Log (wage) difference
	Non-White	White	Non-White	White	
Demographic characteristics					
London	1,100	5,176	2.89	3.10	-0.21***
South East	463	7,060	2.80	2.87	-0.07***
Northern	401	8,716	2.59	2.74	-0.15***
Rest of England	791	14,214	2.64	2.77	-0.13***
Wales	47	4,445	2.54	2.69	-0.15**
Scotland	100	6,140	2.66	2.82	-0.16***
Partnered	1,585	33,462	2.87	2.88	-0.02
Not partnered	1,336	13,273	2.62	2.64	-0.02
With dependent child(ren) in family aged <19	1,366	20,193	2.79	2.90	-0.11***
No dependent child(ren) in family aged <19	1,555	26,542	2.72	2.75	-0.03**
Health problems lasting >1 year	533	11,053	2.72	2.79	-0.07***
No health problems lasting >1 year	2,375	35,526	2.76	2.82	-0.06***
Higher education characteristics					
Russell Group universities	600	11,123	2.93	2.94	-0.01
Non-Russell Group universities	2,321	35,612	2.71	2.77	-0.07***
High status universities	766	15,664	2.89	2.92	-0.03
Other universities	2,155	31,071	2.70	2.76	-0.06***
Good degree	1,640	26,662	2.79	2.82	-0.03**
Other degree class	1,276	20,032	2.71	2.81	-0.10***
Health	251	4,955	2.85	2.82	0.04
Sciences	555	8,844	2.77	2.82	-0.06**
Engineering/Technology	165	4,132	2.89	3.06	-0.17***
Social studies	292	3,901	2.78	2.80	-0.02
Law/Business/Finance	729	6,666	2.73	2.85	-0.12***
Arts/Humanities/Education	417	10,545	2.62	2.68	-0.06**
Combined subject	419	6,480	2.76	2.83	-0.07**
Occupation/Sector characteristics					
Managerial/Professional job	2,041	35,615	2.89	2.94	-0.04***
Other occupation	877	11,100	2.42	2.41	0.00
Public Admin/Education/Health	1,060	20,891	2.71	2.77	-0.05***
Banking/Finance	801	9,239	2.83	2.93	-0.09***
Trade/Hotel/Restaurant	396	4,305	2.47	2.50	-0.03
Transport/Communication	310	4,048	2.90	2.98	-0.09***
Manufacturing/Construction	208	5,046	2.90	2.97	-0.07*
Other industry	139	3,099	2.80	2.75	0.05
Micro/Small enterprises (<50 employees)	1,001	18,010	2.61	2.69	-0.08***
Medium/Large enterprises (>=50 employees)	1,892	28,469	2.83	2.89	-0.06***
Public sector	893	18,195	2.75	2.80	-0.05***
Private sector	2,018	28,401	2.75	2.83	-0.07***
Part-time work	491	8,847	2.52	2.64	-0.12***
Full-time work	2,430	37,882	2.80	2.85	-0.06***
Permanent job	2,717	44,211	2.77	2.83	-0.06***
Non-permanent job	203	2,507	2.51	2.56	-0.05

Note: I examine the equality of wage means between the two groups using standard Student's t-tests. N. Ireland figures are not reported due to the small cell sizes in the non-White category.
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*
Source: APS 2013-2018

Table 4.A5. Economic activity by ethnic group and gender

Economic activity		Men				
		White	**Asian**	**Black**	**Mixed/ Other**	**Total**
In employment	Observations	33,176	1,626	360	457	35,619
	%	97.03	94.70	90.68	92.70	96.80
ILO unemployed	Observations	1,015	91	37	36	1,179
	%	2.97	5.30	9.32	7.30	3.20
Total	**Observations**	**34,191**	**1,717**	**397**	**493**	**36,798**
	%	**100**	**100**	**100**	**100**	**100**
Economic activity		**Women**				
		White	**Asian**	**Black**	**Mixed/ Other**	**Total**
In employment	Observations	37,603	1,619	576	594	40,392
	%	97.77	93.64	94.89	94.74	97.51
ILO unemployed	Observations	859	110	31	33	1,033
	%	2.23	6.36	5.11	5.26	2.49
Total	**Observations**	**38,462**	**1,729**	**607**	**627**	**41,425**
	%	**100**	**100**	**100**	**100**	**100**

Note: This table's sample comprises UK-born first-degree holders (aged 19-65) who are economically active (i.e., employed or unemployed). The International Labour Organisation (ILO) unemployment rate is defined as the share of ILO unemployed people to the total active population (i.e., it excludes inactive people and individuals under 16). The sample size of people in employment in this table is greater than that reported in the descriptive analysis section because it also includes self-employees, people who work but do not disclose wages, and the total wage distribution (that is, it includes the top and bottom 1% of the log(wage) distribution, which was dropped from the initial sample).
Source: APS 2013-2018

Table 4.A6. Employment status by ethnic group and gender

Employment status		Men				
		White	**Asian**	**Black**	**Mixed/ Other**	**Total**
Employee	Observations	28,024	1,380	309	380	30,093
	%	84.52	84.98	85.83	83.15	84.54
Self-employed	Observations	5,132	244	51	77	5,504
	%	15.48	15.02	14.17	16.85	15.46
Total	**Observations**	**33,156**	**1,624**	**360**	**457**	**35,597**
	%	**100**	**100**	**100**	**100**	**100**
Employment status		**Women**				
		White	**Asian**	**Black**	**Mixed/ Other**	**Total**
Employee	Observations	33,561	1,483	527	520	36,091
	%	89.28	91.66	91.49	87.54	89.38
Self-employed	Observations	4,031	135	49	74	4,289
	%	10.72	8.34	8.51	12.46	10.62
Total	**Observations**	**37,592**	**1,618**	**576**	**594**	**40,380**
	%	**100**	**100**	**100**	**100**	**100**

Note: This table's sample comprises UK-born first-degree holders (aged 19-65) who are in employment. The employees' sample size in this table is greater than that reported in the descriptive analysis section because it consists of the total wage distribution (including the top and bottom 1% of the log(wage) distribution, which was dropped from the initial sample), and people who work but do not disclose wages.
Source: APS 2013-2018

Table 4.A7. OLS vs. Quantile regressions: Men
Dependent variable: log(wage)

	Men			
Variable	(1) OLS	(2) 10th quantile	(3) 50th quantile	(4) 90th quantile
Black	-0.183***	-0.194***	-0.171***	-0.124***
	(0.029)	(0.021)	(0.047)	(0.039)
Asian	-0.042***	-0.072***	-0.070***	0.019
	(0.014)	(0.011)	(0.018)	(0.017)
Mixed/Other	-0.013	-0.013	-0.013	-0.019
	(0.023)	(0.012)	(0.031)	(0.024)
Age	0.071***	0.056***	0.071***	0.090***
	(0.002)	(0.002)	(0.002)	(0.003)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
London	0.267***	0.256***	0.273***	0.267***
	(0.009)	(0.009)	(0.010)	(0.010)
South East	0.087***	0.066***	0.087***	0.116***
	(0.009)	(0.010)	(0.009)	(0.012)
Northern	-0.020**	-0.025***	-0.024***	-0.013
	(0.008)	(0.008)	(0.008)	(0.012)
Wales	-0.077***	-0.067***	-0.066***	-0.069***
	(0.010)	(0.014)	(0.011)	(0.015)
Scotland	0.004	0.016	-0.006	0.026*
	(0.009)	(0.011)	(0.009)	(0.014)
N. Ireland	-0.093***	-0.063***	-0.105***	-0.088***
	(0.020)	(0.010)	(0.022)	(0.026)
Partnered	0.085***	0.066***	0.088***	0.096***
	(0.007)	(0.007)	(0.007)	(0.009)
Any dependent children in family aged <19	0.053***	0.055***	0.061***	0.046***
	(0.007)	(0.007)	(0.007)	(0.009)
Health problems lasting >1 year	-0.032***	-0.050***	-0.026***	-0.033***
	(0.007)	(0.008)	(0.007)	(0.009)
Health	0.054***	0.040***	0.039**	0.129***
	(0.017)	(0.012)	(0.016)	(0.026)
Sciences	-0.010	-0.009	-0.015	-0.027*
	(0.011)	(0.011)	(0.012)	(0.015)
Engineering/Technology	0.057***	0.070***	0.050***	0.025
	(0.013)	(0.014)	(0.013)	(0.015)
Law/Business/Finance	0.032***	0.018	0.029**	0.023
	(0.012)	(0.012)	(0.013)	(0.015)
Arts/Humanities/Education	-0.113***	-0.105***	-0.115***	-0.126***
	(0.012)	(0.009)	(0.012)	(0.016)
Combined subject	-0.004	-0.004	-0.007	-0.027
	(0.012)	(0.014)	(0.013)	(0.016)
High-status universities	0.100***	0.077***	0.107***	0.119***
	(0.006)	(0.006)	(0.006)	(0.008)
Good degree	0.063***	0.049***	0.071***	0.079***
	(0.006)	(0.006)	(0.006)	(0.007)
Managerial/Professional job	0.321***	0.290***	0.318***	0.293***
	(0.007)	(0.006)	(0.007)	(0.008)
Public Admin/Education/Health	-0.011	0.047***	-0.017	-0.089***
	(0.013)	(0.011)	(0.014)	(0.015)
Banking/Finance	0.092***	0.096***	0.098***	0.041***
	(0.012)	(0.012)	(0.014)	(0.014)
Trade/Hotel/Restaurant	-0.047***	-0.043***	-0.049***	-0.094***
	(0.014)	(0.013)	(0.015)	(0.019)

Continued on next page

Table 4.A7. (continued)				
Transport/Communication	0.101***	0.102***	0.114***	0.049***
	(0.013)	(0.014)	(0.014)	(0.014)
Manufacturing/Construction	0.081***	0.137***	0.081***	-0.011
	(0.013)	(0.012)	(0.014)	(0.016)
Medium/Large enterprises (>=50 employees)	0.153***	0.171***	0.130***	0.102***
	(0.006)	(0.008)	(0.006)	(0.008)
Public sector	-0.072***	-0.016*	-0.086***	-0.120***
	(0.009)	(0.009)	(0.009)	(0.011)
Part-time work	-0.086***	-0.260***	-0.065***	0.097***
	(0.015)	(0.012)	(0.016)	(0.020)
Permanent job	0.058***	0.113***	0.035**	0.005
	(0.016)	(0.010)	(0.014)	(0.017)
Years in current employer (Tenure)	0.004***	0.009***	0.005***	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)
Total usual weekly hours in main job	0.002***	-0.005***	0.003***	0.007***
	(0.000)	(0.001)	(0.000)	(0.001)
Constant	0.599***	0.754***	0.533***	0.515***
	(0.047)	(0.050)	(0.048)	(0.061)
Survey year/Survey month dummies	Yes	Yes	Yes	Yes
Observations	21,370	21,370	21,370	21,370
Adjusted R^2- Pseudo R^2	0.448	0.248	0.303	0.287

Note: Robust standard errors in parentheses.
** p < 0.1, ** p < 0.05, *** p < 0.01.*
The regression sample comprises UK-born employees (aged 19-65) who hold a first degree from a UK university. It excludes individuals with a higher degree (e.g., Masters, Doctorate).
The base categories for the multi-categorical dummy variables are: “White” (for ethnicity), “Rest of England” (for the region of workplace), “Social studies” (for the subject area of first degree) and “Other sector” (for the industry sector).
Source: APS 2013-2018

Table 4.A8. OLS vs. Quantile regressions: Women

Dependent variable: log(wage)

	Women			
Variable	(1) OLS	(2) 10th quantile	(3) 50th quantile	(4) 90th quantile
Black	-0.046**	-0.041*	-0.035	-0.103***
	(0.018)	(0.021)	(0.024)	(0.012)
Asian	-0.020	-0.025**	-0.009	-0.020
	(0.012)	(0.011)	(0.011)	(0.019)
Mixed/Other	-0.036*	0.014	-0.048***	-0.004
	(0.018)	(0.040)	(0.016)	(0.035)
Age	0.062***	0.049***	0.057***	0.079***
	(0.002)	(0.003)	(0.002)	(0.002)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
London	0.244***	0.233***	0.237***	0.264***
	(0.009)	(0.013)	(0.009)	(0.010)
South East	0.060***	0.052***	0.060***	0.076***
	(0.007)	(0.013)	(0.007)	(0.011)
Northern	-0.001	0.037***	-0.004	-0.022***
	(0.006)	(0.009)	(0.007)	(0.008)
Wales	-0.036***	-0.018	-0.034***	-0.045***
	(0.008)	(0.013)	(0.008)	(0.013)
Scotland	0.041***	0.072***	0.045***	0.016*
	(0.007)	(0.012)	(0.007)	(0.009)
N. Ireland	-0.065***	-0.013	-0.046***	-0.099***
	(0.015)	(0.021)	(0.015)	(0.009)
Partnered	0.056***	0.048***	0.049***	0.072***
	(0.005)	(0.008)	(0.005)	(0.006)
Any dependent children in family aged <19	-0.011*	-0.049***	0.000	0.022***
	(0.005)	(0.009)	(0.006)	(0.007)
Health problems lasting >1 year	-0.025***	-0.032***	-0.019***	-0.035***
	(0.005)	(0.008)	(0.005)	(0.007)
Health	0.021**	0.038***	0.003	0.009
	(0.009)	(0.014)	(0.008)	(0.011)
Sciences	-0.016*	-0.023*	-0.023**	-0.005
	(0.009)	(0.013)	(0.009)	(0.011)
Engineering/Technology	0.018	0.014	0.016	0.041*
	(0.017)	(0.030)	(0.020)	(0.022)
Law/Business/Finance	0.055***	0.029**	0.043***	0.099***
	(0.009)	(0.014)	(0.009)	(0.013)
Arts/Humanities/Education	-0.023***	-0.036***	-0.015*	-0.005
	(0.008)	(0.013)	(0.008)	(0.010)
Combined subject	0.011	-0.002	0.005	0.016
	(0.009)	(0.014)	(0.010)	(0.010)
High-status universities	0.098***	0.069***	0.087***	0.137***
	(0.005)	(0.008)	(0.005)	(0.008)
Good degree	0.035***	0.023***	0.035***	0.028***
	(0.005)	(0.007)	(0.005)	(0.006)
Managerial/Professional job	0.349***	0.299***	0.374***	0.339***
	(0.006)	(0.008)	(0.006)	(0.008)
Public Admin/Education/Health	-0.012	0.043**	-0.011	-0.072***
	(0.011)	(0.020)	(0.012)	(0.021)
Banking/Finance	0.101***	0.107***	0.090***	0.090***
	(0.012)	(0.021)	(0.013)	(0.021)
Trade/Hotel/Restaurant	-0.062***	-0.051**	-0.078***	-0.035
	(0.013)	(0.022)	(0.014)	(0.025)

Continued on next page

Table 4.A8. (continued)				
Transport/Communication	0.095***	0.119***	0.086***	0.098***
	(0.015)	(0.026)	(0.016)	(0.023)
Manufacturing/Construction	0.127***	0.146***	0.102***	0.112***
	(0.014)	(0.024)	(0.017)	(0.023)
Medium/Large enterprises (>=50 employees)	0.078***	0.073***	0.071***	0.056***
	(0.005)	(0.007)	(0.005)	(0.006)
Public sector	0.012*	0.069***	0.009	-0.021**
	(0.006)	(0.010)	(0.007)	(0.009)
Part-time work	-0.011	-0.066***	-0.004	0.036***
	(0.008)	(0.013)	(0.008)	(0.010)
Permanent job	0.022**	0.081***	0.023**	-0.056***
	(0.010)	(0.017)	(0.011)	(0.015)
Years in current employer (Tenure)	0.009***	0.013***	0.010***	0.004***
	(0.000)	(0.001)	(0.000)	(0.001)
Total usual weekly hours in main job	0.003***	0.000	0.004***	0.004***
	(0.000)	(0.001)	(0.000)	(0.000)
Constant	0.771***	0.736***	0.834***	0.845***
	(0.038)	(0.060)	(0.038)	(0.052)
Survey year/Survey month dummies	Yes	Yes	Yes	Yes
Observations	25,900	25,900	25,900	25,900
Adjusted R^2- Pseudo R^2	0.417	0.229	0.283	0.249

Note: Robust standard errors in parentheses.
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*
The regression sample comprises UK-born employees (aged 19-65) who hold a first degree from a UK university. It excludes individuals with a higher degree (e.g., Masters, Doctorate).
The base categories for the multi-categorical dummy variables are: "White" (for ethnicity), "Rest of England" (for the region of workplace), "Social studies" (for the subject area of first degree) and "Other sector" (for the industry sector).
Source: APS 2013-2018

Table 4.A9. Robustness analysis – OLS regressions for employees aged 30-50: by gender

Dependent variable: log(wage)

	Men			Women		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.234***	-0.252***	-0.252***	-0.066***	-0.079***	-0.082***
	(0.039)	(0.038)	(0.036)	(0.025)	(0.027)	(0.024)
Asian	-0.062***	-0.079***	-0.077***	-0.044***	-0.055***	-0.040**
	(0.018)	(0.019)	(0.017)	(0.017)	(0.018)	(0.016)
Mixed/Other	-0.042	-0.074**	-0.052*	-0.058**	-0.085***	-0.069***
	(0.031)	(0.033)	(0.027)	(0.026)	(0.029)	(0.023)
Age	0.064***	0.073***	0.077***	0.073***	0.087***	0.089***
	(0.009)	(0.010)	(0.008)	(0.007)	(0.008)	(0.007)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
London	0.263***	0.275***	0.263***	0.267***	0.285***	0.267***
	(0.012)	(0.013)	(0.011)	(0.012)	(0.013)	(0.010)
South East	0.095***	0.098***	0.081***	0.057***	0.054***	0.054***
	(0.011)	(0.012)	(0.010)	(0.010)	(0.011)	(0.009)
Northern	-0.022**	-0.031***	-0.033***	-0.002	-0.002	-0.002
	(0.010)	(0.011)	(0.009)	(0.009)	(0.009)	(0.008)
Wales	-0.081***	-0.092***	-0.068***	-0.040***	-0.043***	-0.020**
	(0.014)	(0.015)	(0.013)	(0.010)	(0.012)	(0.010)
Scotland	-0.010	-0.020	-0.009	0.029***	0.044***	0.052***
	(0.012)	(0.013)	(0.011)	(0.010)	(0.011)	(0.009)
N. Ireland	-0.120***	-0.129***	-0.132***	-0.065***	-0.055**	-0.030
	(0.027)	(0.029)	(0.027)	(0.019)	(0.022)	(0.019)
Partnered	0.072***	0.091***	0.097***	0.063***	0.089***	0.080***
	(0.010)	(0.011)	(0.010)	(0.007)	(0.008)	(0.007)
Any dependent children in family aged <19	0.074***	0.079***	0.067***	-0.011	0.003	0.000
	(0.009)	(0.009)	(0.008)	(0.007)	(0.008)	(0.007)
Health problems lasting >1 year	-0.036***	-0.045***	-0.035***	-0.029***	-0.042***	-0.047***
	(0.009)	(0.010)	(0.008)	(0.007)	(0.008)	(0.006)
Health	0.035	0.071***	0.110***	0.029**	0.110***	0.118***
	(0.023)	(0.025)	(0.020)	(0.012)	(0.013)	(0.011)
Science	-0.025	-0.016	-0.009	-0.007	0.012	0.019*
	(0.015)	(0.016)	(0.013)	(0.012)	(0.014)	(0.011)
Engineering/Technology	0.030*	0.044**	0.044***	0.036	0.070***	0.064***
	(0.017)	(0.017)	(0.015)	(0.022)	(0.023)	(0.019)
Law/Business/Finance	0.024	0.022	0.028*	0.067***	0.054***	0.053***
	(0.016)	(0.017)	(0.015)	(0.012)	(0.014)	(0.012)
Arts/Humanities/ Education	-0.141***	-0.158***	-0.150***	-0.009	-0.012	-0.012
	(0.016)	(0.017)	(0.014)	(0.011)	(0.012)	(0.011)
Combined	-0.011	-0.009	-0.018	0.030**	0.034**	0.037***
	(0.017)	(0.018)	(0.015)	(0.012)	(0.014)	(0.011)
High-status universities	0.106***	0.117***	0.110***	0.104***	0.125***	0.108***
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)
Good degree	0.079***	0.094***	0.085***	0.041***	0.059***	0.048***
	(0.007)	(0.008)	(0.006)	(0.006)	(0.007)	(0.006)
Managerial/Professional Job	0.341***			0.383***		
	(0.010)			(0.008)		
Public Admin/Education/Health	-0.024	-0.001	-0.013	-0.042***	-0.025	0.016
	(0.017)	(0.018)	(0.015)	(0.016)	(0.017)	(0.014)
Banking/Finance	0.096***	0.121***	0.119***	0.105***	0.134***	0.161***
	(0.016)	(0.017)	(0.015)	(0.017)	(0.018)	(0.016)
Trade/Hotel/Restaurant	-0.069***	-0.087***	-0.079***	-0.084***	-0.131***	-0.096***
	(0.019)	(0.021)	(0.019)	(0.020)	(0.021)	(0.019)

Continued on next page

Table 4.A9. (continued)						
Transport/Communication	0.090*** (0.017)	0.116*** (0.018)	0.118*** (0.016)	0.072*** (0.021)	0.112*** (0.022)	0.152*** (0.020)
Manufacturing/ Construction	0.071*** (0.017)	0.088*** (0.017)	0.094*** (0.015)	0.107*** (0.020)	0.127*** (0.021)	0.172*** (0.018)
Medium/Large enterprises (>=50 employees)	0.164*** (0.008)	0.166*** (0.009)	0.165*** (0.008)	0.080*** (0.007)	0.102*** (0.007)	0.112*** (0.006)
Public sector	-0.094*** (0.012)	-0.085*** (0.012)	-0.052*** (0.010)	-0.004 (0.009)	0.037*** (0.009)	0.040*** (0.008)
Part-time work	-0.159*** (0.025)	-0.226*** (0.027)	-0.192*** (0.022)	0.005 (0.010)	0.038*** (0.012)	0.018* (0.010)
Permanent job	0.085*** (0.029)	0.107*** (0.031)	0.123*** (0.022)	0.029* (0.016)	0.084*** (0.018)	0.102*** (0.013)
Years in current employer (Tenure)	0.003*** (0.001)			0.011*** (0.000)		
Total usual weekly hours in main job	0.003*** (0.001)	0.004*** (0.001)	0.004*** (0.000)	0.004*** (0.000)	0.009*** (0.001)	0.007*** (0.000)
Higher degree			0.074*** (0.008)			0.117*** (0.006)
Constant	0.700*** (0.182)	0.671*** (0.190)	0.590*** (0.159)	0.542*** (0.146)	0.234 (0.162)	0.209 (0.135)
Survey year/Survey month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,386	12,409	16,746	14,822	14,841	20,938
Adjusted R^2	0.345	0.286	0.276	0.348	0.206	0.214

Note: Robust standard errors in parentheses.
** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*
Models (1) and (4): The sample is restricted to UK-born employees aged 30-50 who hold a first degree from a UK university. These models include all variables presented in the main OLS regressions (see Table 4.5).
Models (2) and (5): The sample is restricted to UK-born employees aged 30-50 who hold a first degree from a UK university. These models include all variables presented in the main OLS regressions (Table 4.5), except for the Occupation group and "Years in current employers" (as the latter variables are potentially "bad controls").
Models (3) and (6): These are the same as the models (2) and (5), but they additionally include employees who have a higher degree (Masters, Doctorate, "Postgraduate Certificate in Education", and other postgraduate qualifications). The binary variable "Higher degree" captures postgraduate degree holders.
The base categories for the multi-categorical dummy variables are: "White" (for ethnicity), "Rest of England" (for the region of workplace), "Social studies" (for the subject area of first degree) and "Other sector" (for the industry sector).
Source: APS 2013-2018

5. Final remarks

5.1 Research overview

This research was motivated by the heated social debate about ethnic pay inequalities in the UK labour market, the earlier evidence of ethnic disparities in academic performance, and the Government's goals to help ethnic minorities succeed in higher education and have equal opportunities in their career. Given the gaps in the previous literature around specific aspects of these matters, the current thesis attempts to shed more light on the area by focusing on the following targets.

Firstly, utilising individual-level data from the HESA for the academic years 2010/11-2014/15, I examine whether the probability of degree non-completion differs between young undergraduates from various ethnic backgrounds in the UK universities. Unlike studies to date, I pay particular attention to disentangling the dropout reasons associated with voluntary decisions from those related to failing to meet the academic standards. In doing so, I find that all ethnic minority groups (especially the Black undergraduates) have a higher probability of failing their degrees compared to White students. On the other hand, the latter are more likely to quit voluntarily, for example, because of employment or personal reasons. This approach likely clarifies why some earlier works found conflicting results concerning the effect ethnicity has on the likelihood of dropout, as they could not distinguish between voluntary and involuntary withdrawal (National Audit Office, 2007; Vignoles and Powdthavee, 2009).

Secondly, I explore how the probability of achieving a first-class or upper-second class honours (that is, a "good degree") differs among ethnic groups, conditional on students graduating from higher education. I improve the existing literature by exploiting more recent data, enriched with comprehensive information about prior attainment, parents' social class and other characteristics. I also contribute to the literature by estimating interaction effects between ethnicity and students' gender, social class, university type, and previous attainment. I confirm the earlier research findings that all ethnic minorities have substantially lower chances of earning a good degree than White students (Broecke and Nicholls, 2007; Richardson, 2008). Additionally, I provide a detailed mapping of such ethnic discrepancies in academic attainment. For example, I show that the ethnic inequalities in university attainment for the Asian ethnic groups (especially for Bangladeshi students) are more pronounced amongst women than men (conditional on the total pool of observed characteristics). In contrast, the performance gap of Asian graduates (relative to their White peers) decreases as we move from the top to the bottom social class levels. Black students are the least likely to graduate with a good degree (61%-64%), although their ethnic attainment gap is significantly smaller in the Russell Group institutions than the other types of higher education providers (other pre-1992 and post-1992 universities).

Thirdly, using the most recent data from the Annual Population Survey (2013-2018), I investigate whether UK-born university graduates from ethnic minority backgrounds face substantial pay differences in the UK labour market relative to equally qualified White workers. Employing a series of econometric methods and sensitivity checks, I establish that pay differences for ethnic minorities persist (although at different scales) across most subgroups of employees according to their study subject area, university type, degree class, occupation group, workplace size, and age. This suggests that the underlying dynamics (including discrimination and possible unobserved characteristics) that influence ethnic disadvantages make up a multifaceted scarring effect on the British labour market. Interestingly, the ethnic inequalities manifest after the age of 30. This might imply that some unmeasured characteristics (such as economic motivation, lower expectations about the value of obtaining additional skills within the labour market, network effects, and salary negotiation ability/willingness) aggravate the ethnic wage gaps through the career trajectory. Nevertheless, I provide sound evidence that the impact of such unobserved factors would need to be immense to cancel out the effect of ethnicity on wages, thus documenting the existence of ethnic discrimination, particularly against Black and Asian men.

In brief, across all the key outcomes of this thesis (that is, the likelihood of academic failure, probability of graduating with a good degree, and labour market earnings), the picture is shocking for the Black community. However, other ethnic minority groups also encounter significant hindrances to succeeding at university and being equally represented and equitably remunerated in the labour market.

These inequalities should be addressed for various reasons. First and foremost, from an ethical and legal perspective, it goes without saying that all people should be treated fairly and have equal opportunities to succeed in their life, regardless of their race, gender, socio-economic background, or other characteristics. Second, dropping out of university is linked with misspent resources and economic costs to students, universities, and society (Yorke, 1998). Therefore, reducing ethnic inequalities in university outcomes may also positively affect higher education providers and the economy. In a similar vein, previous work estimates an annual benefit of over 1% of the GDP to the UK economy if ethnic minorities become equally represented in the labour market (McGregor-Smith, 2017). The critical question is, what policies would work to reduce (if not eliminate) the ethnic inequalities documented in the present research?

5.2 Policy implications

In this section, I bring together the main suggested policy initiatives discussed in the conclusion parts of the three empirical chapters of this thesis. The first policy recommendation to the Government would be to introduce a specific and measurable goal to eliminate the ethnic gaps in the dropout probability and monitor its progress within the next few years, especially for Black undergraduates who are more likely to fail university than others. This practice has been successful in the past for promoting ethnic minorities' participation in universities. For example, the Government's goal to raise by 20% the number of ethnic minority people accessing higher education by 2020 relative to 2009 (BIS, 2016) has more than been achieved (Bolton, 2021).

On the universities side, higher education providers should increase their efforts to enhance student support and counselling services. Such actions might alleviate issues that undermine students' academic and social integration into university, which is well-known to influence student retention (Tinto, 1975). Universities could also devise training sessions to develop the study skills of entrants with lower educational profiles, as previous achievement is strongly associated with academic performance and dropout likelihood. Other plausible initiatives include tracking undergraduates' progress (e.g., through implementing learning analytics), adjusting the curriculum content and delivery, and providing further financial support for low-income students.

Improving the information campaigns at schools through outreach programmes to align students' pre-entry expectations with their university experience is also essential, especially for students whose parents/guardians have not attended higher education. Many universities already apply the practices mentioned above, but there is room for further improvement (UUK and

NUS, 2019). Moreover, the present work shows that raising the non-White/White academic staff ratio reduces voluntary and involuntary dropout rates substantially. Therefore, encouraging ethnic diversity in academic environments should become beneficial for ethnic minority students, as it could raise their sense of belongingness to the institution, while non-White academics could also act as role models.

To decrease ethnic wage gaps in the labour market, policy actions should remove impediments to the participation of ethnic minorities in higher-paying jobs and sectors. The present research finds that reducing inequalities relating to job characteristics (such as the uneven representation of non-White employees in managerial/professional occupations and their shorter firm tenure) would halve the pay gap to the benefit of ethnic minorities. In this context, the UK Government might improve equality by introducing, for example, quotas on the minimum proportion of non-White employees in each company (Berson, 2016), primarily aiming to balance the ethnic representation in the higher salariat. Such interventions would be more effective if applied early in employees' career, given that the likelihood of holding managerial positions should be associated with being in the job for many years. Furthermore, new policies could reward companies showing good practice regarding the fair treatment of ethnic minorities (for example, through a deduction in income tax). On employers' side, businesses could diversify the ethnic composition of their recruitment staff, as this would probably increase the likelihood of ethnic minorities being hired and the number of non-White people applying for their vacancies (Giuliano, Levine and Leonard, 2009).

5.3 Research limitations

There are some limitations of this research. First, it is not feasible to control for all the factors that influence the academic performance and non-completion probability using administrative datasets. These factors may relate to cultural attitudes and self-motivation, the wrong choice of the subject of study, learning styles, discrimination in teaching support and assessments, and other reasons pertaining to students' social integration into university (Christie, Munro and Fisher, 2004; Yorke and Longden, 2008; Thomas et al., 2017). If these confounding characteristics vary across ethnic groups systematically, the coefficient estimates for each ethnic group would be biased. For example, if the learning styles differ between ethnic groups (Ridley, 2007), omitting this variable would bias the effect of ethnicity on the dropout likelihood and the probability of graduating with a good degree. The extensive information about individual socio-demographic traits, prior attainment and university-related characteristics used in the econometric analysis rules out these factors' effect on academic attainment and dropout rates. Hence, this work directs national policymakers and universities to the unobserved determinants mentioned above to better understand the persistence of ethnic gaps in higher education outcomes. However, the extent and the direction of the effect these unobserved variables have on ethnic minorities' academic performance remain unknown.

The fact that ethnic minorities are more likely to drop out of university prior to degree completion (because of structural factors or unobserved characteristics) suggests that the selection of students who remain to the end of the degree will be different by ethnicity. This selection issue bears some implications regarding the interpretation of the estimated ethnic effects. For example, the findings on the interaction between gender and ethnicity when examining the likelihood of graduating with a good degree contrast with those relating to the probability of academic failure (see Tables 3.7 and 2.7, respectively). Specifically, females from ethnic minority backgrounds are more likely than males to complete their studies. On the contrary, the ethnic gaps (relative to White students) in the probability of achieving a good degree (conditional on graduating) are on average higher for women than for men. Hence, if the lower-performing men quit university before graduation, this would leave a positive selection still in higher education relative to women of the same ethnicity. In particular, non-White men who remain in university would be positively selected on unobservable characteristics (e.g., representing a higher segment of the ability/motivation distribution) relative to non-White females. In this case, the true ethnic gaps of men in the likelihood of earning a good degree would be underestimated.

Moreover, the broad ethnic disaggregation adopted in the paper examining ethnic pay inequalities in the labour market (because of the limited number of observations for some minorities) hides dissimilarities existing within each of the four ethnic groups (White, Black, Asian, Mixed/Other). This should be of greater importance for the heterogeneous Asian ethnic group. For instance, Pakistani and Bangladeshi employees experience the lowest raw wages among all ethnic groups (as shown in Table 4.1), whereas Indian and Chinese people exhibit higher earnings than the White majority group. The causal effect of ethnicity on earnings is compromised by some unmeasured factors, such as the level of labour market attachment, economic motivation, network and negotiation effects, parental socio-economic background, and pre-university characteristics relating to individuals' school quality and neighbourhood conditions. Similarly, in the absence of persuasive instruments in the APS datasets, this analysis does not adjust for the selection into paid employment, which might be an element of the wage differential, especially amongst women. Therefore, although Oster's (2017) partial identification method adopted in this work establishes the existence of ethnic discrimination against some groups of employees (such as Black and Asian men), it is not feasible to explicitly quantify the magnitude of discrimination relative to other unmeasured characteristics.

Parenthetically, it is noteworthy that there has been a (rather philosophical) debate in the literature on whether researchers can infer causality for "immutable" characteristics, such as race or gender. Holland (1986) mentioned that all individuals are "potentially exposable" to the examined causes in experiments involving treatment and control groups. For example, one can establish the causal effect of schooling on pupils' test performance. However, he contended that causal inferences for immutable traits are incoherent. The first reason is related to the timing of the treatment assignment, given that most control variables are determined after these immutable characteristics are conceived, thus introducing post-treatment bias in their effect on the dependent variable. The second reason is that, unlike randomised experiments, it is practically impossible to manipulate immutable traits, and their counterfactual states cannot be defined. Greiner and Rubin (2011) proposed that one could potentially overcome these issues that compromise causality by centring on how the outcome "deciders" (for instance, employers who set the wages) perceive these traits, instead of looking at the actual traits themselves. Hence, they posit that deciders' perceptions of these characteristics are manipulable, thus allowing for possible policy interventions to change the mechanisms that lead to these perceptions (for example, in employment and earnings discrimination). This becomes more meaningful in the present research context, as ethnicity is a social construct rather than an attribute obtained through birth. In contrast, Marcellesi (2013) provides some arguments against those advocating that race is not a cause. He claims that researchers can draw causal inferences when studying the impact of race on wages or other outcomes, particularly on the grounds of racial discrimination.

5.4 Future work

The following list summarises the future research suggestions provided throughout the present thesis.

- Adopt a mixed-method approach by exploiting qualitative and quantitative data and methods to uncover the impact of hard-to-quantify variables (which are not observed in administrative datasets) on the likelihood of academic failure, voluntary dropout and earning a good degree. This approach would require performing qualitative interviews with a representative group of university students and staff across various UK regions. The relevant questionnaires should contain sections related to students' socio-demographic characteristics, learning styles, cultural attitudes and unfulfilled expectations, institutional structures, and racial discrimination.
- Investigate whether there are disproportionate ethnic effects of the COVID-19 pandemic on academic attainment and the likelihood of university non-completion. Early reports show that ethnic minorities experience more detrimental effects of the pandemic on health, earnings, and employment than White people (Lally, 2020; Bracke et al., 2021). These inequalities could be followed through in ethnic minorities' academic performance.
- Evaluate how ethnic pay inequalities amongst UK-born degree holders evolve over time for recent graduate cohorts. Using a single sample, the present research documents that pay differences deteriorate with age. However, it would also be important to look at how these inequalities develop with time. If ethnic gaps decrease with time, it should imply that recent policy and legislation interventions have tackled to some extent the ethnic inequalities in the UK labour market. Addressing this research question would require future data with more extended analysis periods or longitudinal data, such as the Longitudinal Education Outcomes.
- Understand why the ethnic penalties for Asian men established in the present study are more prominent among employees with lower expected wages (such as those graduating from a non-high-status university, holding a lower degree class, having a non-managerial/professional job, and being employed in small/micro enterprises). Comprehending the underlying

mechanisms that drive these disparities for Asian people would necessitate more years of data (which would increase the sample sizes and statistical power) to distinguish between the main ethnic subgroups (Chinese, Indian, Pakistani, and Bangladeshi) that compose the diverse Asian category.

- Explore how the wage equalisation and full representation of ethnic minority graduates in the labour market would improve the UK's economy and welfare. Building on the approach of McGregor-Smith (2017), this would require exploiting data from the APS and the Annual Survey of Hours and Earnings. For instance, one could firstly estimate the potential employment rates and the proportion of ethnic minority graduates within separate subgroups of individuals based on their occupation, industry sector, region, age, gender, and social class, if all ethnic minorities had the same distribution across the workforce as the White major group. Hence, the difference in total salaries calculated using the current employment rates and the salaries using the potential rates if all ethnic groups were equally represented in the labour market should return the overall economic benefit.

References of Chapter 5

Berson, C., 2016. Local labor markets and taste-based discrimination. *IZA Journal of Labor Economics,* 5(1), pp. 1-21.

BIS, 2016. *Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice.* London: BIS (Department for Business Innovation and Skills).

Bolton, P., 2021. *Higher education student numbers.* London: House of Commons (Library), (7857).

Bracke, P., Croxson, K., Leary, J. and Wood, J., 2021. Covid-19 and the UK's BAME communities – an economic perspective. *Financial Conduct Authority* [Online]. Available from: https://www.fca.org.uk/insight/covid-19-and-uk-bame-communities-economic-perspective [Accessed 20 February 2021].

Broecke, S. and Nicholls, T., 2007. *Ethnicity and degree attainment.* London: Department for Education and Skills, (Research Report RW92).

Christie, H., Munro, M. and Fisher, T., 2004. Leaving university early: exploring the differences between continuing and non-continuing students. *Studies in Higher Education,* 29(5), pp. 617-636.

Giuliano, L., Levine, David I. and Leonard, J., 2009. Manager Race and the Race of New Hires. *Journal of Labor Economics,* 27(4), pp. 589-631.

Greiner, D.J. and Rubin, D.B., 2011. Causal Effects of Perceived Immutable Characteristics. *Review of Economics and Statistics,* 93(3), pp. 775-785.

Holland, P.W., 1986. Statistics and Causal Inference. *Journal of the American Statistical Association,* 81(396), pp. 945-960.

Lally, C., 2020. Impact of COVID-19 on different ethnic minority groups. *UK Parliament (Parliamentary Office of Science and Technology)* [Online]. Available from: https://post.parliament.uk/impact-of-covid-19-on-different-ethnic-minority-groups/ [Accessed 20 February 2021].

Marcellesi, A., 2013. Is Race a Cause? *Philosophy of Science,* 80(5), pp. 650-659.

McGregor-Smith, R., 2017. *Race in the workplace: The McGregor-Smith Review.* London: Department for Business Energy & Industrial Strategy.

National Audit Office, 2007. *Staying the Course: the retention of students in higher education.* London: The Stationery Office, (HC 616).

Oster, E., 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics,* 37(2), pp. 187-204.

Richardson, J.T.E., 2008. The attainment of ethnic minority students in UK higher education. *Studies in Higher Education,* 33(1), pp. 33-48.

Ridley, A., 2007. Approaches to learning, age, ethnicity and assessment. Implications for widening participation. *Psychology Teaching Review,* 13(1).

Thomas, L., Hill, M., O'Mahony, J. and Yorke, M., 2017. *Supporting student success: strategies for institutional change. What Works? Student Retention & Success programme.* London: HEFCE - Higher Education Academy - Paul Hamlyn Foundation - Action on Access. Available from: https://www.phf.org.uk/publications/works-student-retention-success-full/ [Accessed 20 December 2020].

Tinto, V., 1975. Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research,* 45, pp. 89-126.

UUK and NUS, 2019. *Black, Asian and minority ethnic student attainment at UK universities: #Closingthegap.* London: Universities UK (UUK) and National Union of Students (NUS). Available from: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2019/bame-student-attainment-uk-universities-closing-the-gap.pdf [Accessed 20 December 2020].

Vignoles, A. and Powdthavee, N., 2009. The Socioeconomic Gap in University Dropouts. *The B.E. Journal of Economic Analysis & Policy,* 9(1).

Yorke, M., 1998. Non-Completion of Full-Time and Sandwich Students in English Higher Education: Costs to the Public Purse, and Some Implications. *Higher Education,* 36(2), pp. 181-194.

Yorke, M. and Longden, B., 2008. *The first year experience of higher education in the UK.* York: Higher Education Academy.